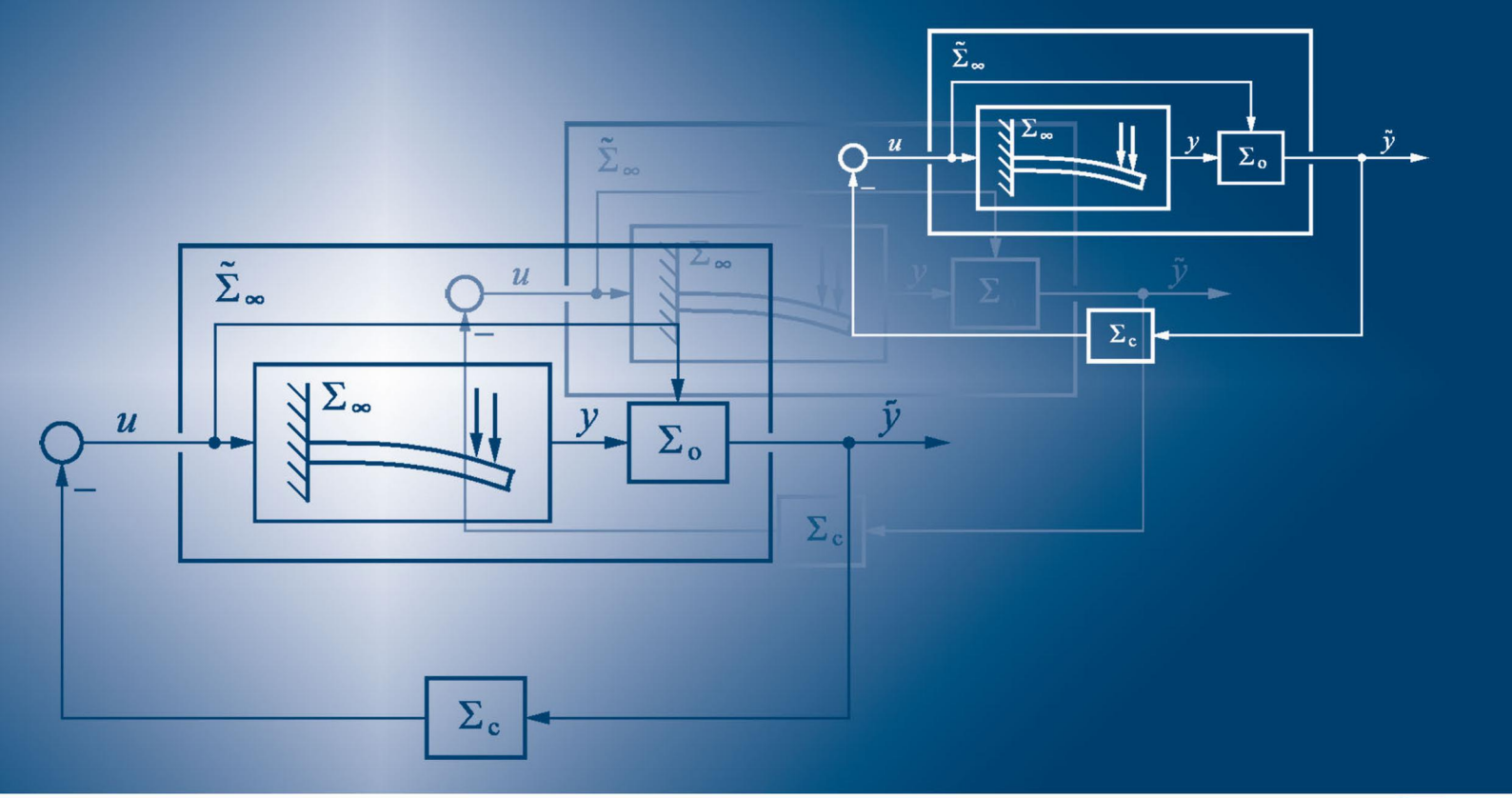

FAU Forschungen, Reihe B, Medizin, Naturwissenschaft, Technik 1

Christian Harkort

Early-Lumping Based Controller Synthesis for Linear Infinite-Dimensional Systems

FAU UNIVERSITY PRESS

Christian Harkort

Early-Lumping Based Controller Synthesis for Linear Infinite-Dimensional Systems

FAU Forschungen, Reihe B

Medizin, Naturwissenschaft, Technik

Band 1

Herausgeber der Reihe:
Wissenschaftlicher Beirat der FAU University Press

Christian Harkort

Early-Lumping Based Controller Synthesis for Linear Infinite-Dimensional Systems

Erlangen
FAU University Press
2014

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Verlag und Auslieferung:

FAU University Press, Universitätsstraße 4, 91054 Erlangen

Druck: docupoint GmbH

ISBN: 978-3-944057-14-9
ISSN: 2198-8102

Early-Lumping Based Controller Synthesis for Linear Infinite-Dimensional Systems

Der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

zur Erlangung des Doktorgrades

DOKTOR-INGENIEUR

vorgelegt von

Dipl.-Ing. Christian Harkort

aus Regensburg

Early-lumping-basierter Reglerentwurf für lineare unendlich-dimensionale Systeme

Der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

zur Erlangung des Doktorgrades

DOKTOR-INGENIEUR

vorgelegt von

Dipl.-Ing. Christian Harkort

aus Regensburg

Als Dissertation genehmigt
von der Technischen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	01.08.2013
Vorsitzende des Promotionsorgans:	Prof. Dr.-Ing. habil. Marion Merklein
Gutachter:	PD Dr.-Ing. habil. Joachim Deutscher
	Prof. Dr. Günter Leugering

Dedicated to my father Heiner.

Danksagung

Die vorliegende Dissertation ist im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Regelungstechnik der Friedrich-Alexander-Universität Erlangen-Nürnberg entstanden.

An erster Stelle möchte ich mich besonders bei meinem Doktorvater Herrn PD Dr.-Ing. habil. Joachim Deutscher bedanken, der als Leiter der Forschungsgruppe "Unendlichdimensionale Systeme" die Promotion mit viel Engagement und Interesse betreut hat. Durch wertvolle Anregungen und konstruktive Kritik in gemeinsamen Diskussionen hat er mich stets unterstützt und damit zum Gelingen der Arbeit beigetragen.

Dem Lehrstuhlinhaber Herrn Prof. Dr.-Ing. habil. Günter Roppenecker möchte ich herzlich dafür danken, dass er mir durchwegs Rückhalt und großen Freiraum zur Bearbeitung der Aufgabenstellung gegeben hat. Darüber hinaus gebührt mein Dank Herrn Prof. Dr. Günter Leugering für die Übernahme des Korreferats und sein detailiertes Interesse an der Arbeit. Ebenso danke ich Herrn Prof. Dr. Christoph Pflaum für die Beteiligung an der Prüfungskommission. Nicht zuletzt möchte ich mich in besonderer Weise bei meinen Kollegen Herrn Dipl.-Ing. Peter Maurer und Herrn Dipl.-Ing. Andreas Mohr für die sorgfältige Durchsicht des Manuskripts sowie für die zahlreichen Anregungen bedanken.

Allen weiteren Kolleginnen und Kollegen des Lehrstuhls möchte ich für die beständige Unterstützung und den Austausch auf fachlicher wie privater Ebene danken. Die Zeit am Lehrstuhl werde ich gerne in guter Erinnerung behalten.

Erlangen, im Mai 2014 Christian Harkort

Abstract

The present thesis is concerned with the finite-dimensional output feedback control of linear time-invariant infinite-dimensional systems with bounded control and measurement. This system class contains a variety of linear parabolic, hyperbolic, and biharmonic distributed-parameter systems with spatially distributed inputs and outputs, as well as some time-delay systems. For these systems an observer-based compensator, *i.e.*, a dynamic output feedback control, is designed by the early-lumping approach. The concept of this approach is to use an approximation of the infinite-dimensional system as the basis for the design of the compensator, which enables to apply the well-known methods for finite-dimensional systems. The impact of the neglected system dynamics on the closed-loop behavior, commonly referred to as spillover, is an inherent shortcoming of the approach that might lead to instability. Usually, this problem is tackled by iterating the compensator design and the closed-loop analysis with a more and more precise approximation until an acceptable control performance is achieved. It is the intention of this thesis to extend the early-lumping-based compensator design method such that the undesired side effect of the model reduction can be diminished more systematically and efficiently. Thereby, approaches both for continuous-time and discrete-time control are presented. In contrast to most contributions dealing with the early-lumping approach the system operator is not assumed to have a discrete spectrum.

In the continuous-time case spillover reduction is achieved by employing fictitious outputs of the infinite-dimensional plant, which are reconstructed by means of so-called output observers of finite order. Since these outputs contain less contributions of the neglected dynamics than the actual outputs of the plant that are available by measurement, the spillover becomes reduced if a fictitious output, instead of the plant output, is used as compensator input. For quantifying the resulting spillover suppression the spectrum of the closed-loop system is analyzed in detail. Under the assumption that the system operator is similar to a normal operator this leads to an estimate for the

spectrum perturbation, that is caused by the spillover. It is shown that this estimate decreases exponentially with respect to the order of the output observers. Additionally, the structure of the closed-loop spectrum is analyzed, leading to the result that this spectrum coincides with the closure of the set of eigenvalues under fairly mild assumptions.

The spillover reduction approach for continuous-time control can also be applied for sampled-data systems in an adapted form, to which end discrete-time output observers have to be added to the control loop. In addition, two further approaches for discrete-time control are presented which require the system operator to be a Riesz-spectral operator. Their central idea is to utilize either a specially adapted hold device or sampling device. These are designed such that the contribution of the neglected dynamics to the system output becomes small at the sampling time instances. The sampled-data system has therefore a transfer behavior that comes arbitrarily close to that of a finite-dimensional system. Thus, the spectrum perturbation can be suppressed to an arbitrary extent, for which an estimate is provided.

The approaches are based on state space representations of the approximation and the neglected dynamics, for which modal approximations are considered.

Zusammenfassung

Diese Arbeit befasst sich mit dem Entwurf endlich-dimensionaler Regelungen für lineare, zeitinvariante unendlich-dimensionale Systeme mit beschränkten Stelleingriffen und Messungen. Diese Systemklasse umfasst eine Vielzahl linearer parabolischer, hyperbolischer und biharmonischer Systeme mit örtlich verteilten Ein- und Ausgängen, sowie eine Klasse von Totzeitsystemen. Die Regler für diese Systeme werden mithilfe des "early-lumping"-Ansatzes entworfen. Bei diesem Verfahren wird das zu regelnde unendlich-dimensionale System durch eine Approximation angenähert, auf deren Basis die Regelung ausgelegt wird. Hierzu können die üblichen Verfahren für endlich-dimensionale Systeme angewendet werden. Die als *Spillover* bezeichnete Auswirkung der vernachlässigten Streckendynamik auf das Regelkreisverhalten stellt eine grundsätzliche Schwäche des Verfahrens dar. Dieser wird üblicherweise dadurch begegnet, dass dem Reglerentwurf eine Analyse der Regelkreisdynamik folgt und der Entwurf ggf. für eine erhöhte Approximationsordnung wiederholt wird, bis sich eine zufriedenstellende Regelkreisgüte einstellt. Ziel der Arbeit ist es, die "early-lumping"-basierte Entwurfsmethode so zu erweitern, dass der unerwünschte Nebeneffekt der Modellvereinfachung systematischer und effektiver reduziert werden kann. Hierfür werden Verfahren sowohl für zeitkontinuierliche als auch für zeitdiskrete Regelungen vorgestellt. Anders als in früheren Arbeiten zum "early-lumping"-Ansatz ist die betrachtete Systemklasse nicht auf Systemoperatoren mit diskretem Spektrum beschränkt.

Im zeitkontinuierlichen Fall wird eine Spillover-Reduzierung durch Verwendung fiktiver Systemausgänge erzielt, welche mittels sogenannter Ausgangsbeobachter endlicher Ordnung gewonnen werden. Da diese Ausgangsgrößen einen geringeren Beitrag der vernachlässigten Dynamik enthalten als die durch Messung verfügbaren Systemausgänge, wird der Spillover-Effekt verringert, wenn ein solcher fiktiver Ausgang anstelle des Streckenausgangs in den Regler eingespeist wird. Zur quantitativen Bewertung der erzielten Spillover-Abschwächung wird eine detaillierte Analyse des Regelkreisspektrums angestellt. Dies führt zu einer Abschätzung der Störung dieses Spektrums,

welche durch den Einfluss der vernachlässigten Dynamik verursacht wird. Als Ergebnis zeigt sich, dass eine obere Schranke der Störung exponentiell bezüglich der Ordnung der Ausgangsbeobachter verringert wird. Dabei ist vorauszusetzen, dass der Systemoperator durch eine Ähnlichkeitstransformation in einen normalen Operator überführt werden kann. Desweiteren zeigt eine Analyse der Struktur des Regelkreisspektrums, dass dieses unter wenig einschränkenden Annahmen dem Abschluss der Menge der Eigenwerte gleicht.

Auch für Abtastregelungen lässt sich das Spillover-Reduktionsverfahren aus dem zeitkontinuierlichen Bereich in angepasster Form anwenden, wozu zeitdiskrete Ausgangsbeobachter in den Regelkreis eingefügt werden. Zusätzlich werden für zeitdiskrete Regelungen zwei weitere Verfahren vorgestellt, wobei der Systemoperator als Riesz-Spektraloperator vorauszusetzen ist. Deren Grundidee besteht darin, entweder ein spezielles Halteglieder oder ein spezielles Abtastglied zu verwenden. Diese werden so entworfen, dass der Beitrag der vernachlässigten Dynamik zum Systemausgang zu den Abtastzeitpunkten vernachlässigbar wird. Das Abtastsystem besitzt dadurch ein Übertragungsverhalten, welches dem eines endlich-dimensionalen Systems beliebig nahegebracht werden kann. Dies hat zur Folge, dass die Störung des Regelkreisspektrums beliebig klein gemacht werden kann. Eine Abschätzung der verbleibenden Störung wird angegeben.

Für die Verfahren werden Zustandsmodelle der Approximation sowie der vernachlässigten Dynamik herangezogen, wobei speziell von modalen Approximationen des unendlich-dimensionalen Systems ausgegangen wird.

Contents

Chapter 1

Introduction

Many processes in the fields of engineering and physics as well as chemistry and biology are dynamical systems whose dynamics depend both on the time and on one or several spatial coordinates. Such systems are commonly called *distributed-parameter systems*. In contrast to their counterpart, the *lumped-parameter systems*, whose behavior can be described by a finite number of scalar variables, the system variables of distributed-parameter systems are elements of an infinite-dimensional function space. These systems belong therefore to the so-called class of *infinite-dimensional systems* which in addition covers also the class of *time-delay systems*. This thesis is concerned with the finite-dimensional output feedback control of linear time-invariant infinite-dimensional systems with bounded control action and measurement. An introduction to the description of such systems is given, *e.g.*, in [46, 51, 67].

The eigenvalue assignment problem for the control of linear distributed-parameter systems by dynamic output feedback has been a wide field of research during the last decades. The restriction to finite-dimensional control, that is essential for the application because infinite-dimensional control laws cannot be implemented directly, is a particular challenging issue. An overview of finite-dimensional control approaches for linear distributed-parameter systems can be found in [5, 37, 42, 51].

It is a widely spread engineering practice to design a compensator, *i.e.*, a dynamic output feedback control, on the basis of a system approximation. At first glance, this so-called *early-lumping approach* seems to be the easiest way of a compensator design for infinite-dimensional systems because a variety of well-developed approximation methods and their software implementations are available, and once the approximation has been computed one can proceed with the established design techniques for lumped-

parameter systems. Of course, since a part of the system dynamics is neglected for the compensator design, the control loop may have an unintended behavior. This negative impact of the neglected system dynamics, which even may destabilize the closed-loop system, is known as *spillover* and was studied intensively during the 1970s and the 1980s (see [3, 4, 5, 8, 105]).

A self-evident possibility to deal with the spillover is to reduce the approximation error sufficiently by increasing the approximation order. In this way the spillover can always be made marginal. However, besides the shortcoming that the resulting compensator order may be undesired high, this approach, though being rather simple from a conceptual point of view, is not easy to deal with in theory in a satisfying way. In particular, an a priori bound for the required compensator order is not available and the design procedure may require several iteration steps consisting of computing a more accurate approximation and redesigning the compensator. The overall method can therefore not be regarded systematic. Although some approaches have been suggested for reducing the spillover (see [13, 31, 105]) the mentioned general difficulties have not been overcome yet.

This is the reason why new concepts such as the *late-lumping approach* (see [11, 12, 42, 46]), the *direct approach* (see [38, 50, 121, 122]) and the robustness-based design approaches (see [23, 42, 46]) have been developed during the 1990s. It is the concept of the late-lumping approach to design an infinite-dimensional compensator first and to approximate it for the implementation in a second step, where the reduction causes spillover similar to the early-lumping approach. The other two approaches, in contrast, avoid the spillover problem at heart by designing the compensator finite-dimensional right from the beginning so that no approximation is involved. Nevertheless, the early-lumping approach still plays an important role for the engineering practice which certainly comes from the fact that its application does not require a deep theoretical background.

It is therefore the intention of this thesis to adhere to the basic idea of early-lumping—to design a compensator on the basis of a finite-dimensional model instead of the infinite-dimensional plant—but to modify the underlaying finite-dimensional model such that the impact of spillover is reduced. This leads to the system analytic problem to identify and characterize the influence of spillover, and to the synthesis problem to extend the classical early-lumping approach such that an efficient and systematic spillover reduction is achieved. The considerations in the thesis for addressing these

problems make use of some basic concepts of functional analysis and the Hilbert space theory, more specifically. However, the bare application of the proposed approaches does not require a deep insight into this field.

1.1 Problem formulation and contribution of the thesis

This thesis is concerned with the design of finite-dimensional compensators for infinite-dimensional linear time-invariant systems on the basis of the early-lumping approach, where both continuous-time and discrete-time compensators are addressed. The general objective is to extend the classical approach such that the spillover phenomenon can be handled more systematically, whereupon a low order of the controller dynamics is desired. For this aim different approaches are developed in this thesis, one for continuous-time control that can be adapted in a straightforward manner to sampled-data control, and two alternative approaches for spillover reduction in discrete-time domain. These techniques can be summarized as follows.

Spillover reduction approach for continuous-time control

The basic degree of freedom, that is used for spillover reduction in the continuous-time domain, consists in combining the infinite-dimensional plant with a suitable finite-dimensional dynamic extension and to design an observer-based compensator for the resulting extended system. In this way, the actual control synthesis can still be done in the classical way but only the system to be controlled is modified. For determining a suitable extension the fact is used that certain fictitious outputs of the infinite-dimensional system can be reproduced by means of a finite-dimensional dynamical system that is named *output observer* (see [56]). These fictitious outputs have the advantage compared to the plant outputs available by measurement, that they are less affected by the neglected dynamics. In consequence, the effect of spillover is reduced when such reconstructed outputs are used for the compensator instead of the measurable outputs. This basic idea can be employed repeatedly by accomplishing the output reconstruction multiple times, whereby the detrimental impact of the system reduction is successively diminished. A class of reconstructible fictitious outputs is identified and a method for the reproduction is given. These results have been published in [55, 56, 77]. A compensator design approach that makes use of a fictitious output is presented subsequently.

In order to end up with a systematic approach it is necessary to analyze the spillover reduction in a quantitative way. To this end, the spectrum of the closed-loop system operator is examined, that determines the behavior of the control loop essentially, as is well-known. Studies on the perturbation of the closed-loop spectrum can be found in [40, 68]. However, in both references the assumptions are too restrictive for the purpose of this thesis and the results do hardly give any advice for spillover avoidance. In this work a simple estimate for the perturbation of the closed-loop spectral points is found that is used as a measure for the spillover effect. On the basis of this result it is shown that by using a fictitious output for the compensator the spectrum perturbation of the control loop is reduced exponentially with respect to the order of the output observers. Therefore, this spillover suppression method turns out to be more effective than the classical practice to increase simply the approximation order which may lead to a comparatively high compensator order. Finally, an a priori estimate for the overall order of the controller dynamics is given that guarantees certain performance criteria. This concept and the results have been presented in [55, 76, 80, 81].

Spillover reduction approach for discrete-time control

It is not surprising that the spillover reduction method for continuous-time control can be applied in an analog way when the early-lumping approach is applied to sampled-data system. This yields also in the discrete-time domain an efficient suppression of the spectrum perturbation. Since the adaption of the approach for continuous-time control to the discrete-time case is straightforward, it is not discussed in the thesis. Instead, two alternative methods are presented that make use of the observation that sampled-data systems provide degrees of freedom for spillover reduction that are unavailable in the domain of continuous-time control. These are related to the choice of the hold device and sampling device that are added to the plant in order to constitute a sampled-data system. The commonly used *zero-order hold*, that keeps the last control vector of the compensator constant during the current sampling interval, is replaced by a more general type of hold device. Its output equals in the single-input case a certain *hold function* which is weighted by the control value that the compensator has generated at the last sampling time instant. Since the hold function is a step function, it can be implemented easily, for which a zero-order hold can be used that operates with a higher sampling rate than the compensator. Alternatively, the commonly used sampling device, that simply takes the plant output vector at the sampling time instances and passes them to the compensator, is replaced by a more general sampling device.

It takes the system output at several time instances within the current sampling interval and computes the weighted sum of these measurements. At each sampling time instance the resulting sum is passed to the discrete-time compensator.

For suppressing the spillover the observation is used that the hold function and the sampling function can be chosen such that the error between the output of the approximation and the infinite-dimensional plant at the sampling time instances becomes arbitrarily small. Thus, the spillover can be reduced without increasing the approximation order or the compensator order, where, in doing so, the implementation of the general hold device or general sampling requires additional efforts, compared to the use of a zero-order hold and standard sampling. The closed-loop dynamics is analyzed as in the continuous-time case and an estimate for the spectrum perturbation is provided. These concepts are addressed in [78, 79].

Generalization of the system class

It is a further objective of the thesis to generalize the early-lumping approach to systems with accumulation points in their spectra. This is motivated by the fact that mechanical structures with Kelvin-Voigt damping, which make up a relevant system class in the applications, possess such accumulation points (see [73, 101]). This generalization impacts particularly the spectral analysis. Almost all contributions in the literature concerning the early-lumping approach consider systems whose spectra consist solely of isolated eigenvalues. These systems do not have any accumulation points since if the eigenvalues accumulate, the closed-loop spectrum contains additional spectral points that are not eigenvalues. Since all kinds of spectral points are responsible for the system's stability it is essential to include also the additional spectral points in the analysis. Therefore, the examination of the closed-loop spectrum's structure is a main issue of the thesis which leads to a detailed characterization.

1.2 Outline of the thesis

Chapter 2 provides in the first section some fundamentals for applying the classical early-lumping approach. Since the considerations in this thesis refer to state space representations of infinite-dimensional systems some basics of such models are summarized. Afterwards, the considered system class is defined and the state space models

for some applications are determined that are used in subsequent chapters. Modal approximations are employed throughout the thesis for the system reduction. For that reason this type of approximation is reviewed in the second section of Chapter 2. Subsequently, the classical finite-dimensional compensator design by the early-lumping approach is summarized in the third section, for which observer-based compensators are considered. Finally, the closed-loop spectrum is analyzed in the last section of Chapter 2 which leads to a characterization of the spillover impact on the controlled system.

The new finite-dimensional compensator design approach is presented in Chapter 3 that combines the classical early-lumping design scheme for observer-based continuous-time control with the proposed spillover reduction technique. To this end, the reconstruction of fictitious outputs by means of output observers is demonstrated in the first section. Afterwards, the actual compensator design approach is addressed in the second section which makes use of fictitious outputs. Both closed-loop structures with a single output observer and with a cascaded setup of output observers are considered. In the last section of Chapter 3 it is shown that the spillover reduction approach enables to suppress two different kinds of spillover, namely the *observation spillover* and the *control spillover*.

Chapter 4 is concerned with the design of digitally implemented compensators that thus operate in the discrete-time domain. In the first section of this chapter some basics of infinite-dimensional sampled-data systems are summarized, and the system class considered for discrete-time control is defined. The classical early-lumping based compensator design approach for such discrete-time systems is discussed subsequently in the second section, wherein also the closed-loop dynamics is analyzed in terms of the spectrum structure and spectrum perturbation.

Finally, the spillover reduction approaches for discrete-time control systems are presented in Chapter 5. While the use of general hold devices is subject of the first section, the second section makes use of general sampling. Besides presenting design methods for these devices estimates for the closed-loop spectrum perturbation are given.

Chapter 2

The early-lumping approach for continuous-time control

The design of compensators for linear distributed-parameter systems on the basis of a finite-dimensional approximation is a classical technique. This so-called *early-lumping approach* enables to use an arbitrary synthesis method for finite-dimensional systems which is why this approach is still wide spread. The early-lumping approach has been studied intensively during the 1970s and the 1980s (see the overview in [5] and [3, 4, 8, 105]).

On the one hand reducing the infinite-dimensional system to a finite-dimensional approximation has the obvious merit that the design step is particular simple. On the other hand this method entails the shortcoming that the neglected system dynamics may affect the behavior of the closed-loop system in an undesired way, which even may lead to instability of the closed-loop system. This phenomenon, known as *spillover*, is analyzed, *e.g.*, in [2, 14, 105]. Since the spectrum of the closed-loop system operator is an essential characteristic of the controlled system with regard to the stability and control performance, it is reasonable to analyze the spillover impact in terms of this spectrum. It turns out that all eigenvalues are shifted away from the desired locations on the complex plane, due to the neglected system dynamics. This perturbation of the spectrum has been studied in [40] and [68]. Unfortunately, the characterization in [40] is fairly involved and does therefore hardly give any advice for a stabilizing compensator, and the latter reference is restricted to special applications and yields estimates that are in many cases very conservative.

For that reason the early-lumping approach is reviewed in this chapter and an estimate

for the spectrum perturbation, *i.e.*, an upper bound for the deviations of the closed-loop eigenvalues caused by spillover, is presented. It is shown that this estimate is proportional to the norm of a perturbation operator that is related to the neglected system dynamics. The estimate allows not only to assure the closed-loop stability but also to guarantee performance criteria such as a minimum stability margin and damping.

Another reason why the early-lumping approach is analyzed from a new perspective is that most of the contributions in this field confine to system operators with discrete spectra. These do therefore not apply to systems with accumulations in the spectrum, such as, *e.g.*, mechanical structures with Kelvin-Voigt damping (see [73, 101]). When the system operator is restricted to have a discrete spectrum the closed-loop system operator's spectrum again is discrete and thus consists solely of isolated eigenvalues. In the general case, however, the structure of the spectrum, *i.e.*, the decomposition into the *point spectrum*, which is the set of eigenvalues, the *continuous spectrum*, and the *residual spectrum* is more involved. The behavior of the controlled system depends on the spectral points of all three categories for which reason it is essential not only to consider the eigenvalues but to identify also the remaining spectral points. Therefore, the decomposition of the closed-loop spectrum is analyzed. As a main result of this chapter it is shown that for a fairly large system class the whole spectrum of the closed-loop system operator always equals the closure of the set of eigenvalues. As a basic conclusion it is sufficient to concentrate on the eigenvalues in order to analyze the closed-loop behavior.

The discussion of the effects of the neglected system dynamics become particular simple when modal approximations are used for the compensator design because the infinite-dimensional system dynamics separates then into two decoupled subsystems in a natural way, namely the modal approximation and the decoupled *residual dynamics*. Therefore, the early-lumping approach using this type of approximation is presented in this chapter.

The chapter is organized as follows. In Section 2.1 some fundamentals of infinite-dimensional state space models are reviewed and the considered system class is defined. In particular, systems with bounded input and output operators are assumed which simplifies the considerations significantly. In addition, some examples are considered which will be referred to repeatedly later on. The computation of modal approximations is summarized in Section 2.2, and Section 2.3 addresses the classical

finite-dimensional observer-based compensator design by the early-lumping approach. Finally, the resulting spillover is discussed by analyzing the closed-loop spectrum in Section 2.4.

2.1 Continuous-time state space representations

The compensator design and the analysis of the closed-loop behavior will be presented throughout the thesis for a certain class of state space models, namely for the class of *state linear systems*. This fairly well-known representation of infinite-dimensional systems is reviewed in the following subsection. Afterwards, important properties, such as, *e.g.*, the solvability of the state equation and the stability of the system are addressed in the Subsections 2.1.2 and 2.1.3. In Subsection 2.1.4 some important subclasses of state linear systems are discussed and the system class, that is considered for the time-continuous control, is defined. Finally, the section ends with some examples of state space models for standard applications.

2.1.1 State linear systems

Distributed-parameter systems are often described by partial differential equations (PDEs), where the describing variables depend on both the time and the spatial coordinates (see, *e.g.*, [102]). These models have therefore a different form compared to the state space representations

$$\dot{x}_n(t) = Ax_n(t) + Bu(t), \qquad t > 0, \quad x_n(0) = x_{n,0} \in \mathbb{C}^n \tag{2.1}$$

$$y(t) = Cx_n(t), \qquad t \geq 0 \tag{2.2}$$

with $u(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^m$, and $x_n(t) \in \mathbb{C}^n$ of finite-dimensional linear systems whose state depends only on the time t. The analogy to these systems can be recovered for infinite-dimensional systems by using a suitable function space as *state space* X, which is the linear space that the state belongs to for all $t \geq 0$, instead of $\mathbb{C}^n$ that is commonly used in finite dimensions. A simple example of such a function space is the *Lebesgue space* $L_2(a, b)$ of complex-valued functions $f : [a, b] \mapsto \mathbb{C}$ that satisfy $\int_a^b |f(z)|^2 \mathrm{d}z < \infty$, wherein $[a, b]$ describes the spatial domain of the system. The state $x(t)$ represents then for every time instant a function that depends on the spatial coordinate $z \in [a, b]$. However, from the set-theoretic point of view, $x(t)$ is just an element of the state

space, *i.e.*, $x(t) \in X$, $\forall\, t \geq 0$, for which reason the spatial argument is omitted in the notation. Therefore, the linear time invariant (LTI) infinite-dimensional systems considered throughout this thesis have the form

$$\Sigma: \quad \dot{x}(t) = \mathcal{A}x(t) + \mathcal{B}u(t), \qquad t > 0, \quad x(0) = x_0 \in X \tag{2.3}$$

$$y(t) = \mathcal{C}x(t), \qquad t \geq 0 \tag{2.4}$$

which apparently is formally almost the same as in the finite-dimensional case. Of course, instead of the matrices A, B, and C in (2.1)–(2.2) the operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in (2.3)–(2.4) are more general linear maps. The *system operator* $\mathcal{A}$ commonly contains spatial differential operators that correspond to the spatial partial derivatives appearing in the models on the basis of PDEs. Therefore, $\mathcal{A}$ can be applied only to functions that are sufficiently smooth which means that the *domain*[1] $D(\mathcal{A})$ of $\mathcal{A}$ is a subset of X. For the systems to be considered the *input operator* $\mathcal{B}$ maps the (possibly vector-valued) input $u(t) \in \mathbb{R}^p$, $p \in \mathbb{N}$, onto a function $\mathcal{B}u(t) \in X$ and the *output operator* $\mathcal{C}$ maps the state onto the (possibly vector-valued) output $y(t) \in \mathbb{R}^m$, $m \in \mathbb{N}$, typically by means of an integral operator. Besides the set-theoretic point of view from which $x(t)$ is an element of X, the geometric perspective plays an important role for generalizing to the infinite-dimensional case. As for finite-dimensional state space models the elements of X and thus $x(t)$ appear from that point of view to be vectors that can be assigned a length by a norm $\|\cdot\|_X$. In addition, it is useful for many considerations to describe the relation between two vectors by aid of an *inner product* $\langle g, h\rangle_X$ with $g, h \in X$. Therefore, the state spaces considered in the sequel consist not only of a function space but of a (complex) *Hilbert space*[2] for which an inner product and the *induced norm* $\|h\|_X := \sqrt{\langle h, h\rangle_X}$, $h \in X$, are defined. At first glance, the model (2.3)–(2.4) does not contain any boundary conditions that are part of PDE-based models. However, the boundary conditions are embedded in the definition of the domain $D(\mathcal{A})$. This is demonstrated in the following simple example, where a state space model of a one-dimensional heat conductor is determined.

[1] The *domain* $D(\mathcal{M})$ of an operator $\mathcal{M} : D(\mathcal{M}) \subseteq H_1 \mapsto H_2$ consists of all elements of H_1 for that the application of $\mathcal{M}$ is defined.

[2] A linear space H over $\mathbb{C}$, in which an inner product $\langle\cdot,\cdot\rangle_H$ is defined, is called a complex *Hilbert space* if it is a *Banach space* with respect to the induced norm $\|h\|_H = \sqrt{\langle h, h\rangle_H}$, $\forall\, h \in H$, *i.e.*, every Cauchy sequence in H has its limit w.r.t. to $\|\cdot\|_H$ in H.

Example 2.1-1 (1-D heat conductor)

The system to be considered is a heat conducting rod with length ℓ, heat conductivity μ, effective heat capacity c and density ρ (see Figure 1). The temperature $\vartheta(z,t)$ along the spatial coordinate $z \in [0,\ell]$ is described by the *heat equation*

$$\partial_t \vartheta(z,t) = \frac{\mu}{c\,\rho}\partial_z^2 \vartheta(z,t) + \frac{1}{c\,\rho} b(z)u(t), \qquad t>0,\ z\in(0,\ell) \tag{2.5}$$

with the initial condition

$$\vartheta(z,0) = \vartheta_0(z), \qquad z\in[0,\ell] \tag{2.6}$$

(see [91]), wherein the abbreviation $\partial_\eta^\alpha = \frac{\partial^\alpha}{\partial \eta^\alpha}$ is used. At both ends of the rod *Dirichlet boundary conditions*

$$\vartheta(0,t) = \vartheta(\ell,t) = 0, \qquad t>0 \tag{2.7}$$

are considered. The input $u(t)$ represents the heating power of a heat source with a spatial distribution that is described by $b : [0,\ell] \mapsto \mathbb{R}$. Finally, a temperature sensor provides the weighted average temperature

$$y(t) = \int_0^\ell \vartheta(z,t)c(z)\mathrm{d}z \tag{2.8}$$

with weight $c : [0,\ell] \mapsto \mathbb{R}$ which is used as the output of the system. If the temperature $\vartheta(z,t)$ is used as the state $x(t) := \vartheta(\cdot,t)$, the system (2.5)–(2.8) can be described by a state space model (2.3)–(2.4) on the complex state space $X := L_2(0,\ell)$ (for the definition of $L_2(0,\ell)$ see Appendix D) which is known to be a Hilbert space with the inner product

$$\langle g,h\rangle_X = \langle g,h\rangle_{L_2} := \int_0^\ell g(z)\overline{h(z)}\mathrm{d}z \tag{2.9}$$

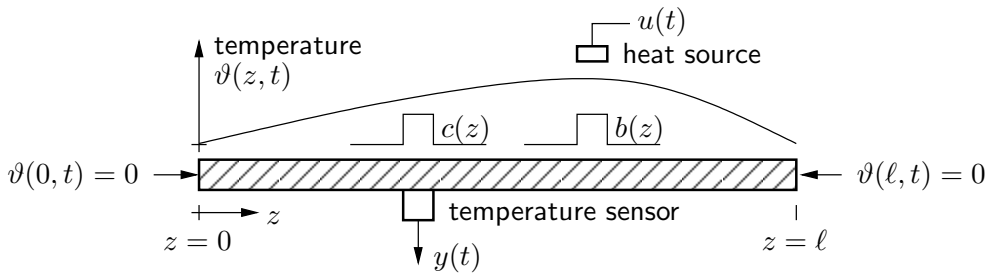

Figure 1 – One-dimensional heat conductor with temperature distribution $\vartheta(z,t)$. The input $u(t)$ is the power of a heat source, and the output $y(t)$ is the average temperature at the sensor surface.

which induces the norm $\|h\|_X = \|h\|_{L_2} := \sqrt{\int_0^\ell |h(z)|^2 \mathrm{d}z}$. It is easy to verify that the corresponding operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are then given by

$$\mathcal{A}h = \frac{\mu}{c\,\rho}\mathrm{d}_z^2 h, \qquad \forall\, h \in D(\mathcal{A}) = \left\{ h \in H_2(0,\ell) \;\middle|\; h(0) = h(\ell) = 0 \right\} \tag{2.10}$$

$$\mathcal{B}v = \frac{1}{c\,\rho} b\, v, \qquad \forall\, v \in \mathbb{C} \tag{2.11}$$

$$\mathcal{C}h = \langle h, c \rangle_X, \qquad \forall\, h \in X, \tag{2.12}$$

in which $\mathrm{d}_\eta^\alpha = \frac{\mathrm{d}^\alpha}{\mathrm{d}\eta^\alpha}$ is used as shorthand notation (for the definition of the Sobolev space $H_2(0,\ell)$ see Appendix D). It is important to note that the domain $D(\mathcal{A})$ reflects the boundary conditions (2.7). In addition, the requirement $h \in H_2(0,\ell)$ within the definition of $D(\mathcal{A})$ assures that $\mathcal{A}$ is applied only to functions that are twice differentiable w.r.t. z in the weak sense. ◀

2.1.2 Solution of the state equation

The state space model (2.3)–(2.4) is meaningful only, if the *abstract initial value problem* (abstract IVP) composed of the state equation in combination with the initial condition (see (2.3)) is *well-posed*, *i.e.*, it has a unique solution which depends continuously on the initial state and on the input trajectory. In order to discuss this question the homogeneous case, *i.e.*, $u \equiv 0$, is considered for the time being. Then, well-posedness of the abstract IVP is ensured if the *semigroup*[3] $\mathcal{S}(t)$, $t \geq 0$, which is the map defined by

$$x(t) = \mathcal{S}(t)x_0, \qquad \forall\, t \geq 0,\ \forall\, x_0 \in X, \tag{2.13}$$

has the property to be *strongly continuous*, *i.e.*, $\lim_{t\to 0+} \|\mathcal{S}(t)x_0 - x_0\|_X = 0,\ \forall\, x_0 \in X$. In this case $\mathcal{S}(t)$ is said to be a C_0-*semigroup* and $\mathcal{A}$ is the corresponding *infinitesimal generator of the* C_0-*semigroup*. Thus, $\mathcal{A}$ will be assumed to be such a generator throughout this thesis so that the homogeneous abstract IVP is well-posed. In order to make sure that the same is valid also for the inhomogeneous abstract IVP the following assumption is used.

[3] A *one-parameter semigroup* or just *semigroup* is a map $\mathcal{S} : \mathbb{R}_+ \mapsto L(X)$ that satisfies $\mathcal{S}(t_1 + t_2) = \mathcal{S}(t_1)\mathcal{S}(t_2)$, $\forall\, t_1, t_2 \geq 0$, where $L(X)$ is the space of bounded maps on X.

Assumption 2.1-2 (Boundedness of the input and output operators)

The operators $\mathcal{B} : \mathbb{C}^p \mapsto X$ and $\mathcal{C} : X \mapsto \mathbb{C}^m$ are assumed to be linear and bounded[4] so that they have the representations

$$\mathcal{B}v = \begin{bmatrix} b_1 & \cdots & b_p \end{bmatrix} v = \sum_{i=1}^{p} b_i v_i, \qquad \forall\, v \in \mathbb{C}^p \tag{2.14}$$

$$\mathcal{C}h = \begin{bmatrix} \langle h, c_1 \rangle_X \\ \vdots \\ \langle h, c_m \rangle_X \end{bmatrix}, \qquad \forall\, h \in X \tag{2.15}$$

with the *input distribution functions* $b_1, b_2, \ldots, b_p \in X$ and the *output distribution functions* $c_1, c_2, \ldots, c_m \in X$ (for (2.15) see [92, Thm. 3.8-1]). Furthermore, $b_1, b_2, \ldots, b_p$ are assumed to be linear independent, and the same is assumed for $c_1, c_2, \ldots, c_m$. ◀

This assumption combined with the property of $\mathcal{A}$ to be an infinitesimal generator of a C_0-semigroup has the consequence that the considered systems belong to the class of *state linear systems* (see [46]). It is well known that the abstract IVP is therefore well-posed not only in the homogeneous but also in the inhomogeneous case. To be more precise, (2.3) has a unique *mild solution* $x(t)$, $t \geq 0$, that can be shown to be continuous with respect to t for every input trajectory $u \in L_q([0, \tau]; \mathbb{R}^p)$, $\tau > 0$, $q \in \mathbb{N}$ (see [46, Lem. 3.1.5]; for the definition of L_q see Appendix D). For the heat conductor considered in Example 2.1-1 Assumption 2.1-2 is discussed next.

Example 2.1-3 (1-D heat conductor, continued)

If the input distribution function $b(z)$ in Example 2.1-1 is non-vanishing on an interval of the z-axis, the heat power enters the system in a spatially distributed way which is why this type of input is called *distributed control*. Similar, if the output distribution function $c(z)$ is non-vanishing on an interval of the z-axis, the sensor action is distributed which is referred to as *distributed measurement*. Clearly, when $b(z)$ and $c(z)$ are real-valued bounded functions on $[0, \ell]$, one has $b, c \in X = L_2(0, \ell)$. Since the the input and output operators $\mathcal{B}$ and $\mathcal{C}$ according to (2.11)–(2.12) have the form (2.14)–(2.15), Assumption 2.1-2 is satisfied in this case. In some applications the size

[4] A linear operator $\mathcal{M} : D(\mathcal{M}) = H_1 \mapsto H_2$ between two normed spaces H_1 and H_2 with norm $\|\cdot\|_{H_1}$ and $\|\cdot\|_{H_2}$, respectively, is called *bounded* if a constant $C \in \mathbb{R}$ exists such that $\|\mathcal{M}h\|_{H_2} \leq C\|h\|_{H_1}$, $\forall\, h \in D(\mathcal{M})$, holds.

of the heat source and the sensor are so small compared to the length ℓ of the rod that it is natural to model them as they would act only at a single point. The heat source is then called a *point actuator* at position $\beta \in [0, \ell]$ and the sensor provides a *point measurement* at position $\gamma \in [0, \ell]$. In fact, this situation can be described by setting $b(z) = \delta_\beta(z)$ and $c(z) = \delta_\gamma(z)$, with $\delta_\zeta(z)$ denoting the *Dirac delta function*[5] centered at ζ, since then the heat enters the system only at the point β and the heat power amounts to $\int_0^\ell b(z)\mathrm{d}z\, u(t) = \int_0^\ell \delta_\beta(z)\mathrm{d}z\, u(t) = u(t)$, and the output becomes $y(t) = \int_0^\ell \vartheta(z,t)c(z)\mathrm{d}z = \int_0^\ell \vartheta(z,t)\delta_\gamma(z)\mathrm{d}z = \vartheta(\gamma,t)$ as intended. In the same way also point actuations and measurements at the boundaries—so-called *boundary control* and *boundary measurement*—can be described. However, since $\delta_\zeta(z) \notin L_2(0,\ell)$, $\zeta \in [0,\ell]$, Assumption 2.1-2 is violated for this type of input and output distribution functions and it is therefore not possible to describe the dynamics by a state linear system as is assumed for the considerations in this thesis. Instead, the larger class of *regular linear systems* can be used that provides a profound theory for systems with point actuation and point measurement (see [44, 130] and the references therein). This approach however requires advanced efforts for the analysis of the well-posedness (see, *e.g.*, [27]). It is therefore in most cases easier to approximate the delta functions by small rectangular shaped functions $b(z) = \frac{1}{2\varepsilon}\mathbf{1}_{[\beta-\varepsilon,\beta+\varepsilon]}(z)$ and $c(z) = \frac{1}{2\varepsilon}\mathbf{1}_{[\gamma-\varepsilon,\gamma+\varepsilon]}(z)$ with $\varepsilon > 0$, where

$$\mathbf{1}_{[\zeta_1,\zeta_2]}(z) := \begin{cases} 1 & : \quad z \in [\zeta_1, \zeta_2] \\ 0 & : \quad \text{else} \end{cases} \tag{2.16}$$

is the *characteristic function*. In this way it is achieved that Assumption 2.1-2 holds. ◀

An explicit expression for the solution x of (2.3) can be determined if the *eigenvectors* ϕ_i, $i \in \mathbb{N}$, of $\mathcal{A}$ have a certain property. The eigenvectors are defined as the non-trivial solutions of the *eigenvalue-eigenvector equation*

$$\mathcal{A}\phi_i = \lambda_i \phi_i, \qquad i \in \mathbb{N},\ \phi_i \in D(\mathcal{A}) \tag{2.17}$$

with λ_i denoting the *eigenvalue* that corresponds to ϕ_i. For determining the solution x of the state equation it is assumed that $\{\phi_i, i \in \mathbb{N}\}$ constitutes a *Riesz basis*, which is the case in many applications of practical interest. Such a basis is defined as follows (see [46, Def. 2.3.1]).

[5] The *Dirac delta function* $\delta_\zeta(z)$, $z \in \mathbb{R}$, has the defining properties $\delta_\zeta(z) = 0$ for all $z \in \mathbb{R} \setminus \zeta$ and $\int_{-\infty}^{\infty} \delta_\zeta(z)\alpha(z)\mathrm{d}z = \alpha(\zeta)$ for any continuous function $\alpha : \mathbb{R} \mapsto \mathbb{C}$.

Definition 2.1-4 (Riesz basis)

Suppose that a set $\{v_i\}_{i\in\mathbb{N}}$ of vectors $v_i \in X$ satisfies both of the following two conditions:

1. $\overline{\text{span}_{i\in\mathbb{N}}}\{v_i\} = X$
2. There exist constants $m, M > 0$ such that

$$m \sum_{i=1}^{N} |\alpha_i|^2 \leq \| \sum_{i=1}^{N} \alpha_i v_i \|_X^2 \leq M \sum_{i=1}^{N} |\alpha_i|^2 \tag{2.18}$$

is satisfied for arbitrary but square summable $\alpha_i \in \mathbb{C}$, *i.e.*, $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$, and for an arbitrary $N \in \mathbb{N}$, in which m and M have to be independent from N.

Then, $\{v_i, i \in \mathbb{N}\}$ is called a *Riesz basis for* X. ◀

One can show that a Riesz basis satisfies (2.18) also for the limit $N \to \infty$, *i.e.*, one has

$$m \sum_{i=1}^{\infty} |\alpha_i|^2 \leq \| \sum_{i=1}^{\infty} \alpha_i v_i \|_X^2 \leq M \sum_{i=1}^{\infty} |\alpha_i|^2 \tag{2.19}$$

for square summable $\alpha_i \in \mathbb{C}$. Note, that the equalities in (2.19) hold with $m = M = 1$ if $\{v_i, i \in \mathbb{N}\}$ is an orthonormal basis for X since the generalized theorem of Pythagoras can be applied. For a general Riesz basis, however, only the inequalities hold with $m, M > 0$. The left inequality implies that the vectors v_i are linearly independent and the right inequality guarantees that any linear combination is contained in X, provided that $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$ holds. This, combined with $\overline{\text{span}_{i\in\mathbb{N}}}\{v_i\} = X$, assures that any element of X can be represented as a unique linear combination of these vectors. Thus, if the eigenvectors ϕ_i, $i \in \mathbb{N}$, of $\mathcal{A}$ are a Riesz basis, every element $h \in X$ has the expansion

$$h = \sum_{i=1}^{\infty} \langle h, \psi_i \rangle_X \phi_i, \qquad \forall\, h \in X \tag{2.20}$$

(see [46, Lem. 2.3.2]), wherein $\{\psi_i, i \in \mathbb{N}\}$ is the (unique) *biorthonormal sequence* associated with $\{\phi_i, i \in \mathbb{N}\}$, *i.e.*,

$$\langle \phi_i, \psi_j \rangle_X = \delta_{ij} := \begin{cases} 1 & : \quad i = j \\ 0 & : \quad \text{else}\,, \end{cases} \tag{2.21}$$

with δ_{ij} denoting the *Kronecker symbol*[6]. An explicit expression for the solution $x(t)$ of system Σ can be obtained by inserting

$$x(t) = \sum_{i=1}^{\infty} \langle x(t), \psi_i \rangle_X \phi_i \tag{2.22}$$

into (2.3). Using the eigenvalue-eigenvector equation (2.17) and relation (2.21) one can show that the state trajectory $x(t)$ is given by

$$x(t) = \sum_{i=1}^{\infty} e^{\lambda_i t} \langle x_0, \psi_i \rangle_X \phi_i + \sum_{i=1}^{\infty} \sum_{j=1}^{p} \int_0^t e^{\lambda_i (t-\tau)} u_j(\tau) \mathrm{d}\tau \, \langle b_j, \psi_i \rangle_X \phi_i, \quad t \geq 0, \; x_0 \in X \tag{2.23}$$

(see [46, Example 3.1.8] and [51, Sec. 2.2.2]), wherein $u_j(t)$ denotes the j-th element of the vector $u(t)$. Remember that for this expression a state linear system was assumed whose system operator has an eigenvector Riesz basis. Without analyzing the convergence of the infinite sums in detail it is obvious that the eigenvalues λ_i have to satisfy $\lambda_0 := \sup_{i \in \mathbb{N}} \operatorname{Re} \lambda_i < \infty$ because one has divergence for any $t > 0$ otherwise. In fact, this condition is necessary for generators of a C_0-semigroup in general. Observe, that (2.23) is analog to the solution

$$x_n(t) = \sum_{i=1}^{n} e^{\lambda_i t} \langle x_0, w_i \rangle_{\mathbb{C}^n} v_i + \sum_{i=1}^{n} \sum_{j=1}^{p} \int_0^t e^{\lambda_i (t-\tau)} u_j(\tau) \mathrm{d}\tau \langle b_j, w_i \rangle_{\mathbb{C}^n} v_i, \quad t \geq 0, \; x_0 \in \mathbb{C}^n \tag{2.24}$$

of a finite-dimensional system (2.1)–(2.2), wherein v_i and w_i denote the eigenvectors of A and $\overline{A}^T$, respectively, that are related to the eigenvalues λ_i and $\overline{\lambda_i}$, and b_j is the j-th column of B. Comparison with (2.23) shows that the eigenvalues influence in both cases the solution and thus the dynamics in the same way. Particularly, the system's stability depends essentially on the eigenvalues which is addressed more in detail in the next subsection.

[6] If $\{\phi_i, i \in \mathbb{N}\}$ is an *orthonormal sequence*, *i.e.*, $\langle \phi_i, \phi_j \rangle_X = \delta_{ij}$, $i, j \in \mathbb{N}$, the biorthonormal sequence is obviously simply given by $\psi_i = \phi_i$, $i \in \mathbb{N}$. The eigenvectors of $\mathcal{A}$ are in fact orthonormal (after normalization), if $\mathcal{A}$ is self-adjoint or skew-adjoint (see [89, Sec. V 3.5]). If $\{\phi_i, i \in \mathbb{N}\}$ is non-orthonormal but the eigenvalues are solely simple, the corresponding biorthonormal sequence $\{\psi_i, i \in \mathbb{N}\}$ can be determined by computing the eigenvectors ψ_i, $i \in \mathbb{N}$, of the *adjoint operator* $\mathcal{A}^*$ (see Appendix C), where ψ_i corresponds to the eigenvalue $\overline{\lambda_i}$ of $\mathcal{A}^*$ with λ_i denoting the eigenvalue of $\mathcal{A}$ that corresponds to ϕ_i. That these vectors ψ_i, $i \in \mathbb{N}$, infact satisfy (2.21) after suitable scaling follows from $\lambda_i \langle \phi_i, \psi_j \rangle_X = \langle \mathcal{A}\phi_i, \phi_j \rangle_X = \langle \phi_i, \mathcal{A}^* \psi_j \rangle_X = \lambda_j \langle \phi_i, \psi_j \rangle_X$ and $\lambda_i \neq \lambda_j$ for $i \neq j$.

2.1.3 Exponential stability

A state linear system and its C_0-semigroup are said to be *exponentially stable* if its norm decays exponentially w.r.t. the time, which means that

$$\|\mathcal{S}(t)x_0\|_X \leq Ce^{\omega t}\|x_0\|_X, \qquad \forall\, t \geq 0,\ \forall\, x_0 \in X \tag{2.25}$$

has to hold for constants $C \geq 1, \omega < 0$. Since $\mathcal{S}(t)x_0$ has the meaning of the homogeneous solution of the state equation (see (2.13)), the C_0-semigroup can be determined for systems with eigenvector Riesz basis by inserting $u \equiv 0$ into (2.23), yielding

$$x(t) = \mathcal{S}(t)x_0 = \sum_{i=1}^{\infty} e^{\lambda_i t}\langle x_0, \psi_i\rangle_X \phi_i, \qquad t \geq 0,\ x_0 \in X. \tag{2.26}$$

A simple calculation gives

$$\begin{aligned}\|\mathcal{S}(t)x_0\|_X^2 &\leq M\sum_{i=1}^{\infty} |e^{\lambda_i t}\langle x_0, \psi_i\rangle_X|^2 \\ &\leq M|e^{\lambda_0 t}|^2 \sum_{i=1}^{\infty} |\langle x_0, \psi_i\rangle_X|^2 \leq \frac{M}{m}(e^{\lambda_0 t})^2\|x_0\|_X^2,\end{aligned} \tag{2.27}$$

wherein (2.19), (2.22), and $\lambda_0 = \sup_{\lambda_i \in \mathbb{N}} \operatorname{Re}\lambda_i$ have been used. Hence, one has

$$\|\mathcal{S}(t)x_0\|_X \leq \sqrt{\frac{M}{m}} e^{\lambda_0 t}\|x_0\|_X. \tag{2.28}$$

Comparison with (2.25) shows that the system is exponentially stable if $\lambda_0 < 0$. Thus, a necessary condition for stability is that all eigenvalues of $\mathcal{A}$ are located in the left half-plane as it is well-known from the finite-dimensional case, whereas they may not come arbitrarily close to the imaginary axis. This remains true also for system operators without eigenvector Riesz basis. However, in contrast to the finite-dimensional case, the *spectrum* $\sigma(\mathcal{A})$ of $\mathcal{A}$ may in general contain not only the eigenvalues λ_i but also additional spectral points which also influence the system's stability. For discussing this more in detail, $\sigma(\mathcal{A})$ is decomposed into three disjoint parts

$$\sigma(\mathcal{A}) = \sigma_p(\mathcal{A}) \cup \sigma_c(\mathcal{A}) \cup \sigma_r(\mathcal{A}), \tag{2.29}$$

where the *point spectrum* $\sigma_p(\mathcal{A})$ contains all the eigenvalues, and the *continuous spectrum* $\sigma_c(\mathcal{A})$ and the *residual spectrum* $\sigma_r(\mathcal{A})$ contain additional spectral points. All $\lambda \in \sigma(\mathcal{A})$ have in common that the equation

$$(\lambda I - \mathcal{A})g = h \tag{2.30}$$

cannot be solved uniquely for arbitrary $h \in X$. If λ is an eigenvalue, *i.e.*, $\lambda \in \sigma_p(\mathcal{A})$, an eigenvector $\phi \in X$ exists such that $(\lambda I - \mathcal{A})\phi = 0$ holds, implying a loss of the uniqueness of (2.30). This relation shows that $\lambda I - \mathcal{A}$ loses its *injectivity*[7] whenever λ is an eigenvalue of $\mathcal{A}$, which therefore can be considered the basic property of an eigenvalue. Spectral points λ belonging to the residual or the continuous spectrum in contrast are defined by the requirement that the *range*[8] $\operatorname{ran}(\lambda I - \mathcal{A})$ does not cover the whole state space so that (2.30) has no solution for any $h \in X \setminus \operatorname{ran}(\lambda I - \mathcal{A})$. In other words, these spectral points $\lambda \in \sigma_r(\mathcal{A}) \cup \sigma_c(\mathcal{A})$ can be characterized by $\lambda I - \mathcal{A}$ losing the *surjectivity*[9] but being injective. Finally, the difference between $\sigma_r(\mathcal{A})$ and $\sigma_c(\mathcal{A})$ is that for $\lambda \in \sigma_c(\mathcal{A})$ one has $\operatorname{ran}(\lambda I - \mathcal{A})$ *dense*[10] in X which means that an arbitrary small change of $h \in X \setminus \operatorname{ran}(\lambda I - \mathcal{A})$ suffices for (2.30) to be solvable. In contrast, this is not possible for $\lambda \in \sigma_r(\mathcal{A})$ since $\operatorname{ran}(\lambda I - \mathcal{A})$ is then not dense in X (see [92]). For completeness, (2.30) is solvable for arbitrary $h \in X$ if λ belongs to the *resolvent set*

$$\rho(\mathcal{A}) \coloneqq \mathbb{C} \setminus \sigma(\mathcal{A}) \tag{2.31}$$

so that the *resolvent (operator)* $(\lambda I - \mathcal{A})^{-1}$ is a bounded operator that is defined on whole X if $\lambda \in \rho(\mathcal{A})$.

In summary, not only the eigenvalues but all spectral points in $\sigma(\mathcal{A})$ must necessarily have negative real parts and may not come arbitrary close to the imaginary axis for $\mathcal{S}(t)$ to be exponential stable. A quantitative characterization of the behavior of C_0-semigroup is its *growth bound* ω_0, which is the infimum of all real numbers ω for that (2.25) holds. For system operators with eigenvector Riesz basis (2.28) makes apparent that the growth bound is given by $\omega_0 = \lambda_0$. However, the relation

$$\omega_0 = \sup_{\lambda \in \sigma(\mathcal{A})} \operatorname{Re} \lambda \tag{2.32}$$

may not hold in the more general case without eigenvector Riesz basis. If it does hold, $\mathcal{A}$ and $\mathcal{S}(t)$ are said to satisfy the *spectrum determined growth assumption* (SDGA). In view of the above reasoning the following statement is immediate.

---

[7] An operator $\mathcal{M} : D(\mathcal{M}) = H_1 \mapsto H_2$ is said to be *injective* if the implication $\mathcal{M}g = \mathcal{M}h \Rightarrow g = h$, $\forall\, g.h \in H_1$ holds. This is equivalent to the implication $g \neq h \Rightarrow \mathcal{M}g \neq \mathcal{M}h$, $\forall\, g, h \in H_1$.

[8] The *range* $\operatorname{ran}\mathcal{M}$ of an operator $\mathcal{M} : D(\mathcal{M}) \subseteq H_1 \mapsto H_2$ is the image of $D(\mathcal{M})$ under $\mathcal{M}$, *i.e.*, $\operatorname{ran}\mathcal{M} = \{g \in H_2 \,|\, \exists\, h \in D(\mathcal{M}) : g = \mathcal{M}h\}$.

[9] An operator $\mathcal{M} : D(\mathcal{M}) = H_1 \mapsto H_2$ is said to be *surjective* if for any $g \in H_2$ it exists at least one element $h \in D(\mathcal{M})$ such that $\mathcal{M}h = g$.

[10] A subspace V of a Hilbert space H is said to be *dense in* H if its closure $\overline{V}$ satisfies $\overline{V} = H$.

Proposition 2.1-5

Suppose that $\mathcal{A}$ is the generator of a C_0-semigroup. If the eigenvectors of $\mathcal{A}$ constitute a Riesz basis, then $\mathcal{A}$ satisfies the spectrum determined growth assumption.

Since the subsequent chapters discuss the compensator design by the eigenvalue assignment method it is reasonable to assume that $\mathcal{A}$ satisfies the SDGA.

2.1.4 Classes of system operators

The considerations in the subsequent chapters simplify when the system operator is restricted to a certain system class that is established now.

Besides the discussed properties of operators $\mathcal{A}$ with eigenvector Riesz basis these operators have the pleasant representation

$$\mathcal{A}h = \mathcal{A}\left(\sum_{i=1}^{\infty}\langle h, \psi_i\rangle_X \phi_i\right) = \sum_{i=1}^{\infty}\lambda_i\langle h, \psi_i\rangle_X \phi_i, \qquad \forall\, h \in D(\mathcal{A}) \tag{2.33}$$

which is called *modal decomposition* (see (2.17) and (2.20)). An important class of such operators are the *Riesz-spectral operators*. These are particular important because they have some properties that avoid technical difficulties on the one hand and appear in many applications on the other hand. Their definition is given next (see [46, Def. 2.3.4]).

Definition 2.1-6 (Riesz-spectral operator)

Suppose that a linear operator $\mathcal{A} : D(\mathcal{A}) \subset X \mapsto X$, whose domain $D(\mathcal{A})$ is dense in X, has the following properties:

1. $\mathcal{A}$ is closed[11],
2. the eigenvalues λ_i, $i \in \mathbb{N}$, of $\mathcal{A}$ are *isolated*[12] and simple,

[11] An operator $\mathcal{M} : D(\mathcal{M}) \subset H_1 \mapsto H_2$ is said to be *closed* if for every sequence $h_k \in D(\mathcal{M})$ such that h_k converges in H_1 and $\mathcal{M}h_k$ converges in H_2 one has $h := \lim h_k \in D(\mathcal{M})$ and $\lim \mathcal{M}h_k = \mathcal{M}h$.

[12] An element s of a subset S of a metric space (V, d) is called *isolated* if an $\varepsilon > 0$ exists such that $d(s, \tilde{s}) > \varepsilon$ holds for every $\tilde{s} \in S \setminus s$. In particular, an eigenvalue $\lambda_i \in \sigma_p(\mathcal{A})$ is isolated if an $\varepsilon > 0$ exists such that $|\lambda_i - \lambda| > \varepsilon$ holds for every $\lambda \in \sigma_p(\mathcal{A}) \setminus \lambda_i$.

3. $\overline{\sigma_p(\mathcal{A})}$ is *totally disconnected*, *i.e.*, no two points $a, b \in \overline{\sigma_p(\mathcal{A})}$ with $a \neq b$ can be joined by a continuous open curve lying entirely in $\overline{\sigma_p(\mathcal{A})}$[13],

4. the eigenvectors ϕ_i, $i \in \mathbb{N}$, of $\mathcal{A}$ form a Riesz basis for X.

Then, $\mathcal{A}$ is called a *Riesz-spectral operator*. ◀

An important property of these operators is, that although their eigenvalues are isolated they may accumulate. That means that the point spectrum $\sigma_p(\mathcal{A})$ contains convergent sequences, whose limits are called *accumulation points*. Different from the eigenvalues λ_i, each of which has a neighborhood that entirely belongs to the resolvent set $\rho(\mathcal{A})$, the accumulation point does not have such a neighborhood because the eigenvalues come arbitrarily close to it. Therefore, $\sigma_p(\mathcal{A})$ is *discrete*[14] while $\overline{\sigma_p(\mathcal{A})}$ is not. Besides the fact that the eigenvectors of Riesz-spectral operators form a Riesz basis by assumption so that these operators satisfy the SDGA, their spectrum has the properties

$$\sigma(\mathcal{A}) = \overline{\sigma_p(\mathcal{A})} = \overline{\{\lambda_i,\, i \in \mathbb{N}\}} \tag{2.34}$$

$$\sigma_r(\mathcal{A}) = \emptyset \tag{2.35}$$

(see [46, Thm. 2.3.5] and [73, Thm. 2.8]). This is important because it is therefore sufficient to consider the operator's eigenvalues in order to analyze the stability of the generated semigroup $S(t)$ since (2.32) and (2.34) imply that the *eigenvalue criterion*

$$\sup_{i\in\mathbb{N}} \operatorname{Re}\lambda_i < 0 \qquad \Longleftrightarrow \qquad \mathcal{S}(t) \text{ is exponentially stable} \tag{2.36}$$

is valid. A Riesz-spectral operator is the infinitesimal generator of a C_0-semigroup if its eigenvalues satisfy $\sup_{i\in\mathbb{N}} \operatorname{Re}\lambda_i < \infty$. This situation is assured if the *sector condition*

$$\exists\, a \in \mathbb{R}, \varepsilon \in [0, \tfrac{1}{2}\pi) : \quad \sigma_p(\mathcal{A}) \subset S_{a,\varepsilon} \coloneqq \Big\{\lambda \in \mathbb{C} \;\Big|\; |\arg(a-\lambda)| \leq \frac{1}{2}\pi - \varepsilon\Big\}, \tag{2.37}$$

is satisfied, *i.e.*, a sector $S_{a,\varepsilon}$ with $\varepsilon \geq 0$ as depicted in Figure 2 can be found such that it contains all eigenvalues of $\mathcal{A}$. Therefore, the following operator class is important for the considerations in this thesis.

[13] Let M be a set. Then, $\overline{M}$ has the meaning of the closure of M. If M is in contrast a complex number, $\overline{M}$ denotes the complex conjugate of M.

[14] A subset S of a metric space is called *discrete* if each element $s \in S$ is isolated. In particular, $S \subset \mathbb{C}$ is discrete if for every $s \in S$ an $\varepsilon > 0$, depending on s, exists such that $|s - \tilde{s}| > \varepsilon$ holds for every $\tilde{s} \in S \setminus s$.

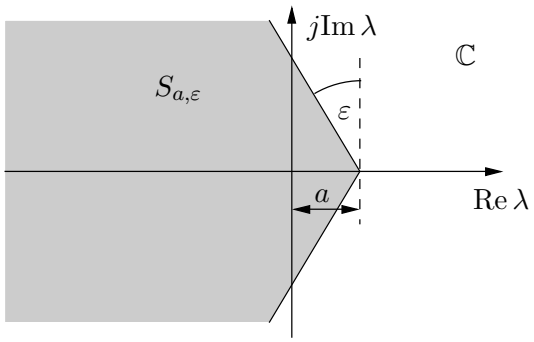

Figure 2 – Sector $S_{a,\varepsilon}$ related to the sector condition.

Definition 2.1-7 (Sectorial Riesz-spectral operator)

Suppose that a linear operator $\mathcal{A} : D(\mathcal{A}) \subset X \mapsto X$, whose domain $D(\mathcal{A})$ is dense in X, has the following properties:

1. The eigenvalues λ_i, $i \in \mathbb{N}$, of $\mathcal{A}$ satisfy the sector condition (2.37),
2. the eigenvalues λ_i, $i \in \mathbb{N}$, are isolated and simple,
3. $\overline{\sigma_p(\mathcal{A})}$ is totally disconnected,
4. the eigenvectors ϕ_i, $i \in \mathbb{N}$, of $\mathcal{A}$ form a Riesz basis for X.

Then, $\mathcal{A}$ is called a *sectorial Riesz-spectral operator*. ◀

Different from Definition 2.1-6 there is no closedness requirement for sectorial Riesz-spectral operators. In fact, this property is implied by the Items 1 and 4 of Definition 2.1-7. If the sector condition holds even for a sector angle $\varepsilon > 0$ it is assured that $\mathcal{A}$ generates a C_0-semigroup that is analytic[15]. This plays an important role in Section 2.3 for the SDGA w.r.t. the closed-loop system. This leads to the following statement that is proven in Appendix A.1.

Proposition 2.1-8

[15] $\mathcal{S}(t)$ is said to be an *analytic C_0-semigroup* if it is a C_0-semigroup that is defined for $t \in \Delta := \{z \in \mathbb{C} : |\arg z| < \theta\} \cup \{0\}$ with $0 < \theta < \pi/2$ and $\mathcal{S}(t)$ is an analytic function w.r.t. $t \in \Delta \setminus \{0\}$. Since $x_h(t) = \mathcal{S}(t)x_0$ is the homogeneous solution of (2.3) this implies that x_h is smooth in the sense of $x_h \in C^\infty(0, \infty)$. An example of analytic C_0-semigroups are those corresponding to heat conduction systems, where the smoothness of the solution can be motivated by the pronounced low-pass character of these systems.

If $\mathcal{A}$ is a sectorial Riesz-spectral operator, then it is a Riesz-spectral operator that is the generator of a C_0-semigroup. Moreover, if the sector condition holds for $\varepsilon > 0$, then the generated C_0-semigroup is analytic.

So far, it has been argued that system operators $\mathcal{A}$ will be considered in the following chapters that generate C_0-semigroups and that satisfy the SDGA. For simplifying the considerations concerning the continuous-time control it turns out to be helpful to impose some additional restrictions on the spectrum $\sigma(\mathcal{A})$. These are summarized in the following assumption that thereby defines the considered operator class.

Assumption 2.1-9 (Properties of the system operator)

$\mathcal{A} : D(\mathcal{A}) \subset X \mapsto X$ is assumed to have the following properties:

1. $\mathcal{A}$ is the infinitesimal generator of a C_0-semigroup.
2. $\mathcal{A}$ satisfies the spectrum determined growth assumption.
3. The spectrum $\sigma(\mathcal{A})$ of $\mathcal{A}$ satisfies all the following conditions:
   a) $\sigma_p(\mathcal{A})$ consists solely of isolated eigenvalues with finite multiplicities,
   b) $\overline{\sigma_p(\mathcal{A})}$ is totally disconnected,
   c) $\sigma_r(\mathcal{A}) = \emptyset$,
   d) $\sigma(\mathcal{A}) = \overline{\sigma_p(\mathcal{A})}$. ◀

While the Items 1 and 2 have been explained already above the Items 3a–3d concerning the spectrum of $\mathcal{A}$ are satisfied by most of the system operators relevant for the applications and are needed to avoid technical difficulties. Particularly, systems with $\mathcal{A}$ being a Riesz-spectral operator and time-delay systems have these properties (see Definition 2.1-6 and [46]). The task to check the items of Assumption 2.1-9 is simplified if $\mathcal{A}$ is a sectorial Riesz-spectral operator by the following statement that immediately follows from the mentioned properties of these operators.

Proposition 2.1-10

If $\mathcal{A}$ is a sectorial Riesz-spectral operator, then $\mathcal{A}$ satisfies Assumption 2.1-9.

In some applications $-\mathcal{A}$ belongs to the more specific class of *Sturm-Liouville operators* which have the following form (see [106, Def. 7.5.1]).

Definition 2.1-11 (Sturm-Liouville operator)

Let $p : [a,b] \mapsto \mathbb{R}$, $q : [a,b] \mapsto \mathbb{R}$, and $\rho : [a,b] \mapsto \mathbb{R}$ be continuous functions such that $\rho(z) > 0$ and $p(z) > 0$ for all $z \in [a,b]$, and $\mathrm{d}p/\mathrm{d}z$ is continuous. Then, a linear operator of the form

$$\mathcal{A}_{SL}h = \frac{1}{\rho(z)}\left(\frac{\mathrm{d}}{\mathrm{d}z}\left(-p(z)\frac{\mathrm{d}h}{\mathrm{d}z}(z)\right) + q(z)h(z)\right) \tag{2.38}$$

$$D(\mathcal{A}_{SL}) = \left\{h \in H_2(a,b) \;\middle|\; \alpha_a\frac{\mathrm{d}h}{\mathrm{d}z}(a) + \beta_a h(a) = 0,\ \alpha_b\frac{\mathrm{d}h}{\mathrm{d}z}(b) + \beta_b h(b) = 0\right\}, \tag{2.39}$$

with $(\alpha_a, \beta_a) \neq (0,0)$ and $(\alpha_b, \beta_b) \neq (0,0)$ is called a *Sturm-Liouville operator*. ◀

It can be shown that $\mathcal{A}$ is a sectorial Riesz-spectral operator if $-\mathcal{A}$ is a Sturm-Liouville operator (see [49, Lem. 1] and [106, Thm. 7.5.6]). Using this makes it particularly simple to check Assumption 2.1-9 by aid of Proposition 2.1-10.

Example 2.1-12 (1-D heat conductor, continued)

For the system operator $\mathcal{A}$ in Example 2.1-1, defined in (2.10), $-\mathcal{A}$ satisfies (2.38)–(2.39) for $a = 0$, $b = \ell$, $\rho(z) \equiv 1$, $p(z) \equiv \frac{\mu}{c\rho}$, $q(z) \equiv 0$, $\alpha_a = 0$, $\beta_a = 1$, $\alpha_b = 0$, and $\beta_b = 1$. Thus, $-\mathcal{A}$ is a Sturm-Liouville operator and hence $\mathcal{A}$ a sectorial Riesz-spectral operator. Therefore, $\mathcal{A}$ satisfies Assumption 2.1-9 according to Proposition 2.1-10. ◀

Next, the state space models for some applications are established and shown to belong to the assumed system class.

2.1.5 Examples of state space models

In this subsection the state space models of a heat conducting plate, an Euler-Bernoulli beam with structural damping, an Euler-Bernoulli beam with Kelvin-Voigt damping, and a time-delay system are determined. These models will be used repeatedly in the subsequent chapters. The first example, examining a heat conducting plate, is in many aspects similar to the one-dimensional heat conductor considered before. The

difference is however, that now two spatial coordinates have to be taken into account.

Example 2.1-13 (Heat conducting plate)

A rectangular heat conducting plate as shown in Figure 3 is considered. The temperature $\vartheta(z_1, z_2, t)$ along the two spatial coordinates $z_1 \in [0, 1]$ and $z_2 \in [0, \pi]$ is described by the state space model

$$\dot{x}(t) = \mathcal{A}x(t) + \mathcal{B}u(t), \qquad t > 0,\ x(0) = x_0 \in X \tag{2.40}$$

$$y(t) = \mathcal{C}x(t), \qquad t \geq 0 \tag{2.41}$$

with the state $x(t) = \vartheta(\cdot, \cdot, t)$ on the state space $X = L_2(\Omega)$ with $\Omega = (0, 1) \times (0, \pi)$, which is known to be a Hilbert space with the usual inner product $\langle u, v\rangle_X = \langle u, v\rangle_{L_2(\Omega)} = \iint_\Omega u(z_1, z_2)\overline{v(z_1, z_2)}\mathrm{d}z_1\mathrm{d}z_2$. The operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are given by

$$\mathcal{A}h = \frac{\partial^2 h}{\partial z_1^2} + \frac{\partial^2 h}{\partial z_2^2}, \qquad \forall\, h \in D(\mathcal{A}) = \left\{ h \in H_2(\Omega) \;\middle|\; h\,\big|_\Gamma = 0 \right\} \tag{2.42}$$

$$\mathcal{B}v = b\,v, \qquad \forall\, v \in \mathbb{C} \tag{2.43}$$

$$\mathcal{C}h = \langle h, c\rangle_X, \qquad \forall\, h \in X \tag{2.44}$$

(see [91]), in which $\Gamma := \partial\Omega$ denotes the boundary of the plate, and $H_2(\Omega)$ is defined in Appendix D. The input distribution function $b(z_1, z_2)$ and the output distribution function $c(z_1, z_2)$ are bounded functions that characterize the distribution of the heat

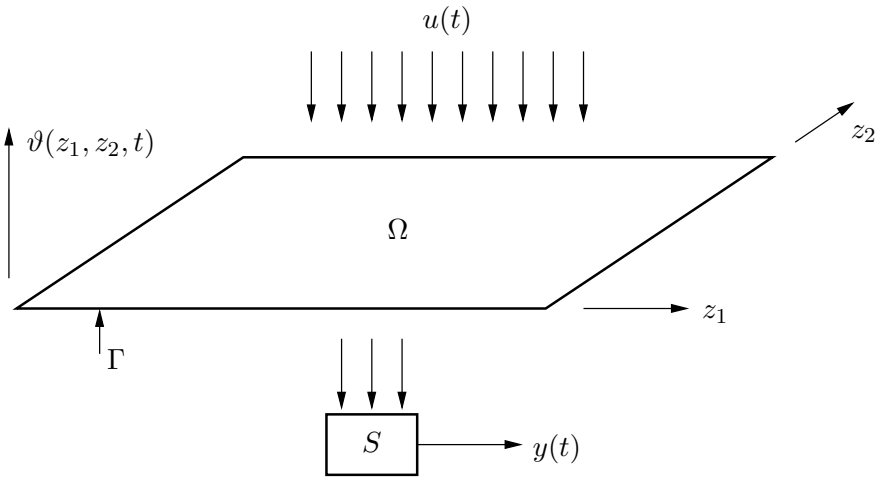

Figure 3 – Heat conducting plate with temperature $\vartheta(z_1, z_2, t)$. The input $u(t)$ is the power of a heat source whose distribution over the plate is described by $b(z_1, z_2)$, and the output $y(t)$ is the weighted average temperature with weight $c(z_1, z_2)$.

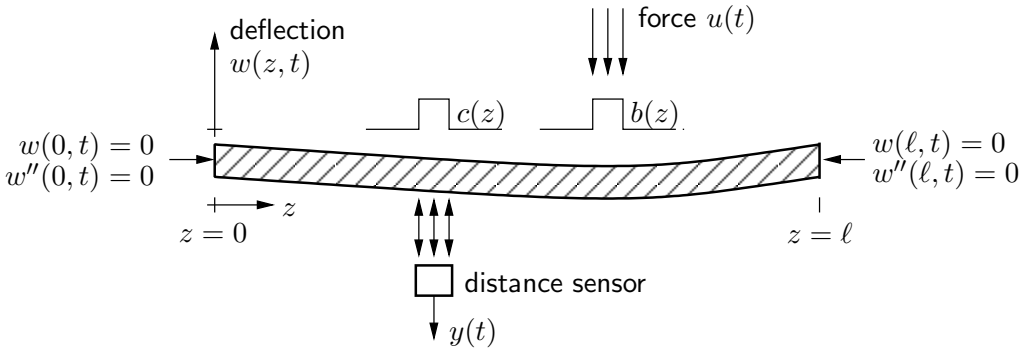

Figure 4 – Euler-Bernoulli beam which is actuated by a force $u(t)$. The resulting displacement $w(z,t)$ is measured by a sensor that provides the average displacement at the sensing area as output $y(t)$.

power u over the spatial domain and the averaging weight for the output, respectively (compare to Example 2.1-1). The system operator $\mathcal{A}$ has the point spectrum

$$\sigma_p(\mathcal{A}) = \{\lambda_{j,k} \,|\, j,k \in \mathbb{N}\} = \{-j^2\pi^2 - k^2 \,|\, j,k \in \mathbb{N}\}. \tag{2.45}$$

The eigenvalues $\lambda_{j,k}$ of $\mathcal{A}$ are isolated and simple, which satisfy the sector condition for, *e.g.*, $a = 0$ and $\varepsilon = \frac{1}{4}\pi$, and $\overline{\sigma_p(\mathcal{A})}$ is totally disconnected. The normalized corresponding eigenvectors are given by

$$\phi_{j,k}(z_1,z_2) = \frac{2}{\sqrt{\pi}} \sin(j\pi z_1) \sin(kz_2), \qquad j,k \in \mathbb{N} \tag{2.46}$$

which form an orthonormal basis and hence a Riesz basis for $X = L_2(\Omega)$. Thus, $\mathcal{A}$ is a sectorial Riesz-spectral operator in view of their Definition 2.1-7 so that Assumption 2.1-9 holds according to Proposition 2.1-10. ◀

In the next example it will be demonstrated that the choice of the state coordinates can be more difficult than it was for the heat conductors in the Examples 2.1-1 and 2.1-13. Different from these examples the state is not scalar-valued but a vector of two spatially distributed variables in the following example.

Example 2.1-14 (Euler-Bernoulli beam with structural damping)

The state space representation of a flexible one-dimensional beam with a single force actuator and a displacement sensor is considered as depicted in Figure 4. The transverse

displacement $w(z,t)$ of the beam along the spatial coordinate $z \in [0,\ell]$ is described by the *Euler-Bernoulli beam model* with structural damping

$$\partial_t^2 w(z,t) = -\partial_z^4 w(z,t) + 2\delta\partial_z^2\partial_t w(z,t) + b(z)u(t), \qquad t>0,\ z\in(0,\ell) \tag{2.47}$$

and the initial condition

$$w(z,0) = w_0(z), \qquad \partial_t w(z,0) = v_0(z), \qquad z\in[0,\ell] \tag{2.48}$$

which is based on the assumptions that cross-sections remain plane under deformation, and the shear strain and the rotational inertia are neglected (see, *e.g.*, [28, 75, 85, 128]). The term $2\delta\partial_z^2\partial_t w(z,t)$ takes *structural damping* into account with the damping constant δ. The input $u(t)$ in (2.47) represents a force acting on the spatial interval $[\beta_1,\beta_2]$ with $0 \le \beta_1 < \beta_2 \le \ell$. The influence of this force on the displacement is therefore described by the input distribution function $b(z) = \frac{1}{\beta_2-\beta_1}\cdot \mathbf{1}_{[\beta_1,\beta_2]}(z)$. The beam is *simply supported*, *i.e.*, the displacement and the momentum at both ends of the beam vanish which is described by the boundary conditions

$$w(0,t) = w(\ell,t) = 0, \qquad t>0. \tag{2.49}$$

and

$$\partial_z^2 w(0,t) = \partial_z^2 w(\ell,t) = 0, \qquad t>0, \tag{2.50}$$

respectively. Finally, the output $y(t)$ of the system is the average displacement within the sensing area of a displacement sensor, *i.e.*,

$$y(t) = \int_0^\ell w(z,t)c(z)\mathrm{d}z, \qquad t\ge 0 \tag{2.51}$$

with the output distribution function $c(z) = \frac{1}{\gamma_2-\gamma_1}\cdot \mathbf{1}_{[\gamma_1,\gamma_2]}(z)$, where $[\gamma_1,\gamma_2]$ with $0 \le \gamma_1 < \gamma_2 \le \ell$ is the interval that the sensor is located on. In what follows, the *energy coordinates* $x_1(t) \coloneqq \partial_z^2 w(\cdot,t)$ and $x_2(t) \coloneqq \partial_t w(\cdot,t)$ are used for a state space model with state vector $x(t) = [x_1(t)\ \ x_2(t)]^T$. The space $X = L_2(0,\ell)\oplus L_2(0,\ell)$ turns out to be a suitable state space, wherein the symbol $\oplus$ denotes the *direct sum*[16], which is a Hilbert space with the inner product

$$\left\langle \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \right\rangle_X = \int_0^\ell g_1(z)\overline{h_1(z)}\mathrm{d}z + \int_0^\ell g_2(z)\overline{h_2(z)}\mathrm{d}z. \tag{2.52}$$

[16] The *direct sum* $V = V_1 \oplus V_2$ of two linear spaces V_1 and V_2 is the set $V_1 \times V_2$ equipped with the addition operation $(v_1,v_2)+(w_1,w_2) = (v_1+w_1, v_2+w_2)$ and the scalar multiplication $c(v_1,v_2) = (cv_1, cv_2)$ for all $(v_1,v_2)\in V_1\times V_2$.

In order to describe the operators of the state space model the operator $\mathcal{A}_0 : D(\mathcal{A}_0) \subset L_2(0,\ell) \mapsto L_2(0,\ell)$ is introduced that is defined by

$$\mathcal{A}_0 h = \mathrm{d}_z^2 h, \qquad \forall\, h \in D(\mathcal{A}_0) = \big\{ h \in H_2(0,\ell) \;\big|\; h(0) = h(\ell) = 0 \big\}. \tag{2.53}$$

Then, it is straightforward to verify that (2.47) can be represented by

$$\dot{x}(t) = \begin{bmatrix} 0 & \mathcal{A}_0 \\ -\mathcal{A}_0 & 2\delta\mathcal{A}_0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ b \end{bmatrix} u(t), \qquad x(0) = x_0 := \begin{bmatrix} \mathrm{d}_z^2 w_0 \\ v_0 \end{bmatrix}. \tag{2.54}$$

Thus, the operators $\mathcal{A}$ and $\mathcal{B}$ in (2.3) are given by

$$\mathcal{A} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 0 & \mathcal{A}_0 \\ -\mathcal{A}_0 & 2\delta\mathcal{A}_0 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} \mathcal{A}_0 h_2 \\ \mathcal{A}_0(-h_1 + 2\delta h_2) \end{bmatrix}, \qquad \forall \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \in D(\mathcal{A}) \tag{2.55}$$

$$\mathcal{B} v = \begin{bmatrix} 0 \\ b \end{bmatrix} v, \qquad \forall\, v \in \mathbb{C}, \tag{2.56}$$

where $D(\mathcal{A}) = D(\mathcal{A}_0) \oplus D(\mathcal{A}_0)$. Note, that the definitions of $D(\mathcal{A})$ and $D(\mathcal{A}_0)$ ensure directly the boundary conditions (2.50) but do not solely imply (2.49). Instead, $\partial_t w(0,t) = \partial_t w(\ell,t) = 0$, $t > 0$, is implied. However, since x_0 always satisfies $x_0(0) = x_0(\ell) = 0$ due to the mechanical setup, all the boundary conditions in (2.49)–(2.50) are assured. It is known that $\mathcal{A}_0$ is self-adjoint with a bounded inverse on $L_2(0,\ell)$ (see [46, Example 2.2.5]) so that $w(t) = \mathcal{A}_0^{-1} x_1(t)$. Hence, (2.51) becomes

$$y = \left\langle \begin{bmatrix} \mathcal{A}_0^{-1} x_1(t) \\ 0 \end{bmatrix}, \begin{bmatrix} c \\ 0 \end{bmatrix} \right\rangle_X = \left\langle \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \begin{bmatrix} \mathcal{A}_0^{-1} c \\ 0 \end{bmatrix} \right\rangle_X, \tag{2.57}$$

where it is used that also $\mathcal{A}_0^{-1}$ is self-adjoint, and this yields

$$\mathcal{C} h = \left\langle h, \begin{bmatrix} \mathcal{A}_0^{-1} c \\ 0 \end{bmatrix} \right\rangle_X = \left\langle h, \begin{bmatrix} \tilde{c} \\ 0 \end{bmatrix} \right\rangle_X, \qquad \forall\, h \in X. \tag{2.58}$$

Therein, $\tilde{c} := \mathcal{A}_0^{-1} c$ is obtained from solving the boundary value problem

$$\mathcal{A}_0 \tilde{c} = \mathrm{d}_z^2 \tilde{c} = c, \qquad \tilde{c}(0) = \tilde{c}(\ell) = 0 \tag{2.59}$$

in view of (2.53) which yields the piecewise defined polynomial

$$\tilde{c}(z) = -\frac{1}{2\ell} \begin{cases} (2\ell - \gamma_1 - \gamma_2) z & : \quad z \in [0, \gamma_1) \\ \dfrac{\ell z^2 + (\gamma_2^2 - \gamma_1^2 - 2\ell\gamma_2) z + \ell\gamma_1^2}{\gamma_1 - \gamma_2} & : \quad z \in [\gamma_1, \gamma_2) \\ -(\gamma_1 + \gamma_2) z + \ell(\gamma_1 + \gamma_2) & : \quad z \in [\gamma_2, \ell]. \end{cases} \tag{2.60}$$

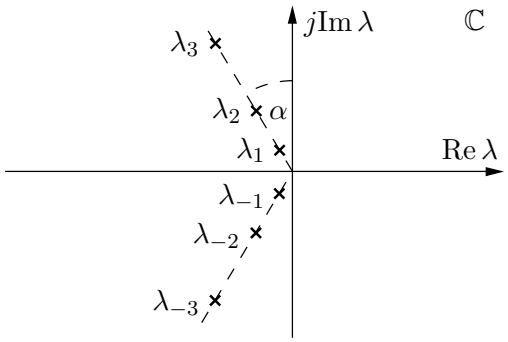

Figure 5 – Eigenvalue distribution of the Euler-Bernoulli beam with structural damping.

Thus, (2.56) and (2.58) show that Assumption 2.1-2 is satisfied. It can be verified easily that the operator $\mathcal{A}$ has the isolated and simple eigenvalues

$$\lambda_{\pm i} = (-\delta \pm j\sqrt{1-\delta^2})\,\frac{\pi^2 i^2}{\ell^2}, \qquad i \in \mathbb{N}, \tag{2.61}$$

which means that the eigenvalues are located on rays through the origin of the complex plane with angle $\alpha = \arcsin(\delta)$ (see Figure 5), where the absolute values $|\lambda_{\pm i}| = \pi^2 i^2$ increase quadratically with respect to i. Of course, one has to make sure that the eigenvalues (2.61) describe whole $\sigma_p(\mathcal{A})$. That this is true is shown in [51, Sec. 2.2.2]. The eigenvalues satisfy the sector condition (2.37) for $a = 0$ and $\varepsilon = \alpha$, and $\overline{\sigma_p(\mathcal{A})}$ is totally disconnected. In [51, Sec. B.2] it is also shown that the eigenvectors corresponding to $\lambda_{\pm i}$, that are given by

$$\phi_{\pm i}(z) = \frac{1}{\sqrt{\ell}} \sin\left(\frac{\pm i\pi z}{\ell}\right) \begin{bmatrix} 1 \\ -\lambda_{\pm i}/|\lambda_{\pm i}| \end{bmatrix}, \qquad i \in \mathbb{N}, \tag{2.62}$$

form a Riesz basis for X. Therefore, $\mathcal{A}$ is a sectorial Riesz-spectral operator so that Assumption 2.1-9 holds due to Proposition 2.1-10. Thus, both Assumptions 2.1-2 and 2.1-9 are satisfied by the state space model (2.54) and (2.57). The choice $x_1(t) = \partial_z^2 w(\cdot, t)$ has the shortcoming that the deflection w is not represented by the state directly. Instead, it is obtained from solving the boundary value problem corresponding to $\mathcal{A}_0 w(\cdot, t) = x_1(t)$.

It should be remarked that the property of the eigenvectors $\phi_{\pm i}$ to be a Riesz basis does not depend on the boundary conditions but instead holds in general (see [28]). In contrast, they do not form a Riesz basis if the more intuitive choice $\tilde{x}(t) = [w(\cdot, t)\ \ \partial_t w(\cdot, t)]^T$ of the state and the inner product as defined above is used for which $\mathcal{A}$ is even not a C_0-semigroup generator. That this choice of the state does not make

sense from a physical point of view can be understood by considering the sequences $w_k(z) = k^{-1}\sin(k\pi z/\ell)$, $\partial_t w_k(z) \equiv 0$, $k \in \mathbb{N}$, of the deflection and the deflection velocity. Then,

$$\|\tilde{x}\|_X^2 = \langle \tilde{x}, \tilde{x}\rangle_X = \frac{1}{2}\frac{\ell}{k^2} \to 0 \qquad \text{for} \quad k \to \infty \tag{2.63}$$

shows that the state converges to its equilibrium. However, the potential energy

$$E_k = \frac{1}{2}\int_0^\ell (\partial_z^2 w(z,t))^2 \mathrm{d}z = \frac{1}{4}\frac{k^2\pi^4}{\ell^3} \tag{2.64}$$

stored in the system increases ad infinitum for $k \to \infty$. Thus, the behavior of the state $\tilde{x}$ does not represent reasonably the actual physical behavior of the beam. In contrast, the state x considered before yields $\|x\|_X^2 = 2E_k \to \infty$, indicating correctly that the system does not come to rest. This shows that the choice of the state coordinates is a critical issue. ◀

It is a strength of the analysis and design approaches presented in the following chapters that systems with accumulation points in their spectra are covered by the considered system class. An example for such a system is given next.

Example 2.1-15 (Euler-Bernoulli beam with Kelvin-Voigt damping)

As in the previous example the Euler-Bernoulli beam depicted in Figure 4 is considered with the only difference that instead of structural damping now *Kelvin-Voigt damping* is assumed which is characteristic for visco-elastic materials such as many polymers. The Euler-Bernoulli beam model with Kelvin-Voigt damping reads

$$\partial_t^2 w(z,t) = -\partial_z^4 w(z,t) - 2\delta\partial_z^4\partial_t w(z,t) + b(z)u(t), \qquad t > 0,\ z \in (0,\ell) \tag{2.65}$$

(see [29, 85, 101]). This type of damping, reflected by the term $-2\delta\partial_z^4\partial_t w(z,t)$, corresponds to the physical effect that the stress is proportional to the rate of strain changes (see [47]). As in the previous example again ℓ is the beam length, $z \in [0,\ell]$ the spatial coordinate, and δ is the damping constant. The beam is simply supported so that the boundary conditions (2.49)–(2.50) hold, and the distributed input and output are described as before by $b(z) = \frac{1}{\beta_2-\beta_1}\mathbf{1}_{[\beta_1,\beta_2]}(z)$ and (2.51) with $c(z) = \frac{1}{\gamma_2-\gamma_1}\mathbf{1}_{[\gamma_1,\gamma_2]}(z)$, respectively. A state space model of the beam with energy coordinates $x(t) := [\partial_z^2 w(\cdot,t)\ \ \partial_t w(\cdot,t)]^T$ on the state space $X := L_2(0,\ell) \oplus L_2(0,\ell)$ with the inner product $\langle \left[\begin{smallmatrix} g_1 \\ g_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} h_1 \\ h_2 \end{smallmatrix}\right]\rangle_X = \int_0^\ell g_1(z)\overline{h_1(z)}\mathrm{d}z + \int_0^\ell g_2(z)\overline{h_2(z)}\mathrm{d}z$ can be determined

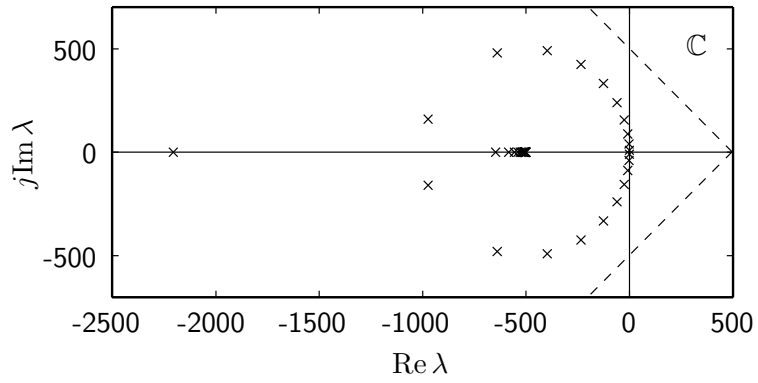

Figure 6 – Distribution of the eigenvalues $\lambda_{\pm i}$ of the Euler-Bernoulli beam with Kelvin-Voigt damping for the damping constant $\delta = 10^{-3}$ and length $\ell = 1$. While the branch λ_{-i} has the asymptotical decrease $\lambda_{-i} \to -2\delta(i\pi)^4$, the eigenvalues λ_i accumulate at $\lambda^{acc} = -500$ for $i \to \infty$. The eigenvalues satisfy the sector condition for the sector bounded by the dashed lines.

in the same way as in Example 2.1-14. The only difference is that $\mathcal{A}$ in (2.55) now becomes

$$\mathcal{A}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} \mathcal{A}_0 h_2 \\ -\mathcal{A}_0(h_1 + 2\delta\mathcal{A}_0 h_2) \end{bmatrix}, \qquad \forall \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \in D(\mathcal{A}) \tag{2.66}$$

$$D(\mathcal{A}) = \left\{ \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \in D(\mathcal{A}_0) \oplus D(\mathcal{A}_0) \;\middle|\; h_1 + 2\delta\mathcal{A}_0 h_2 \in D(\mathcal{A}_0) \right\} \tag{2.67}$$

with $\mathcal{A}_0$ defined in (2.53). The operators $\mathcal{B}$ and $\mathcal{C}$ remain the same as in (2.56) and (2.58) with (2.60) so that Assumption 2.1-2 is satisfied. The main difference between the Euler-Bernoulli beam with structural damping and the Euler-Bernoulli beam with Kelvin-Voigt damping lies in the eigenvalue distribution of $\mathcal{A}$. The eigenvalues

$$\lambda_{\pm i} = \left(-\delta \left(\frac{i\pi}{\ell} \right)^2 \pm \sqrt{\delta^2 \left(\frac{i\pi}{\ell} \right)^4 - 1} \right) \left(\frac{i\pi}{\ell} \right)^2, \qquad i \in \mathbb{N} \tag{2.68}$$

of $\mathcal{A}$, whose locations in the complex plane are depicted in Figure 6, form two branches. While the eigenvalues λ_{-i}, $i \in \mathbb{N}$, with negative index decay asymptotically according to

$$\lim_{i\to\infty} \lambda_{-i} = -2\delta \left(\frac{i\pi}{\ell} \right)^4, \tag{2.69}$$

the eigenvalues λ_i, $i \in \mathbb{N}$, with positive index satisfy

$$\lim_{i\to\infty} \lambda_i = \lambda^{acc} = -\frac{1}{2\delta}. \tag{2.70}$$

Thus, different from the Euler-Bernoulli beam with structural damping, whose eigenvalues are located on rays through the origin of the complex plane (see Figure 5), the beam with Kelvin-Voigt damping has an accumulation point λ^{acc} in its spectrum. Although this spectral point is not isolated because the eigenvalues come arbitrarily close to it, each of the eigenvalues is isolated. The eigenvalues are simple in view of (2.68) if $\delta \neq (\ell/i\pi)^2$ holds for all $i \in \mathbb{N}$. Finally, it is clear that $\overline{\sigma_p(\mathcal{A})}$ is totally disconnected and the sector condition (2.37) holds for, *e.g.*, $a = 500$ and $\varepsilon = \pi/4$ (see Figure 6). The eigenvectors $\phi_{\pm i}$ of $\mathcal{A}$ that correspond to $\lambda_{\pm i}$ are given by

$$\phi_i(z) = \frac{1}{\sqrt{\ell}} \sin\left(\frac{i\pi}{\ell} z\right) \begin{bmatrix} 1 \\ \frac{-\lambda_i \ell^2}{(i\pi)^2} \end{bmatrix}, \qquad i \in \mathbb{N} \tag{2.71}$$

$$\phi_{-i}(z) = \frac{1}{\sqrt{\ell}} \sin\left(\frac{i\pi}{\ell} z\right) \begin{bmatrix} \frac{\ell^2}{(i\pi)^2} \\ \frac{-\lambda_{-i} \ell^4}{(i\pi)^4} \end{bmatrix}, \qquad i \in \mathbb{N} \tag{2.72}$$

and form a Riesz basis in X. In summary, $\mathcal{A}$ is a sectorial Riesz-spectral operator[17], and Assumption 2.1-9 is satisfies due to Proposition 2.1-10. ◀

A type of systems that are infinite-dimensional but do not have spatially distributed parameters are time-delay systems. The initial state x_0 of state space models for such systems consists of a trajectory of lumped-parameter states over a time interval with the duration of the system's time-delays. Since these trajectories belong to a function space it is clear that the state space must be infinite-dimensional. An example for this system class is considered next.

Example 2.1-16 (Time-delay system)

The time-delay system shown in Figure 7 is described by

$$\dot{x}_l(t) = A_l x_l(t) + A_d x_l(t - T) + B_l u(t), \qquad t > 0 \tag{2.73}$$

$$y(t) = C_l x_l(t), \qquad t \geq 0 \tag{2.74}$$

with $x_l(t) \in \mathbb{C}^r$, $u(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^m$, and delay $T > 0$. In order to describe this system by a state space model, the delay term $x_l(t - T)$ in (2.73) can be expressed by

[17] Therefore, $\mathcal{A}$ is a Riesz-spectral operator and thus in particular closed in view of Definition 2.1-6. In fact, the result in [28] stating that $\mathcal{A}$ is not closed is false.

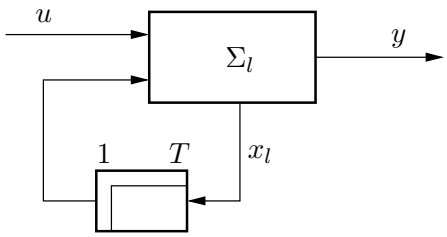

Figure 7 – Structure of a time-delay system with delayed state feedback.

aid of the *transport equation*

$$\partial_t x_d(\tau,t) = \partial_\tau x_d(\tau,t), \qquad t>0,\ \tau\in(-T,0) \tag{2.75}$$

$$x_d(0,t) = x_l(t), \qquad t>0 \tag{2.76}$$

which has the solution

$$x_d(\tau,t) = x_l(t+\tau), \qquad t\geq 0,\ \tau\in[-T,0]. \tag{2.77}$$

Thus, by inserting $x_d(-T,t) = x_l(t-T)$ into (2.73) the system (2.73)–(2.74) can be written as

$$\begin{bmatrix} \dot{x}_l(t) \\ \partial_t x_d(\tau,t) \end{bmatrix} = \begin{bmatrix} A_l & A_d(\cdot)|_{\tau=-T} \\ 0 & \partial_\tau(\cdot) \end{bmatrix} \begin{bmatrix} x_l(t) \\ x_d(\tau,t) \end{bmatrix} + \begin{bmatrix} B_l \\ 0 \end{bmatrix} u(t) \tag{2.78}$$

$$y(t) = \begin{bmatrix} C_l & 0 \end{bmatrix} \begin{bmatrix} x_l(t) \\ x_d(\tau,t) \end{bmatrix}, \tag{2.79}$$

wherein the time-delay does not occur explicitly anymore. Instead, (2.79) is a state space model of the form (2.3)–(2.4) for the state $x(t) \coloneqq [x_l(t)\ \ x_d(\cdot,t)]^T$ with

$$\mathcal{A}\begin{bmatrix} h_l \\ h_d \end{bmatrix} = \begin{bmatrix} A_l h_l + A_d h_d(-T) \\ \mathrm{d}_\tau h_d \end{bmatrix}, \qquad \forall \begin{bmatrix} h_l \\ h_d \end{bmatrix} \in D(\mathcal{A}) \tag{2.80}$$

$$\mathcal{B}v = \begin{bmatrix} B_l \\ 0 \end{bmatrix} v, \qquad \forall\, v\in\mathbb{C} \tag{2.81}$$

$$\mathcal{C}h = \begin{bmatrix} C_l & 0 \end{bmatrix} h, \qquad \forall\, h\in X \tag{2.82}$$

and the state space

$$X \coloneqq \mathbb{C}^r \oplus L_2([-T,0];\mathbb{C}^r) \tag{2.83}$$

(for the definition of $L_2([-T,0];\mathbb{C}^r)$ see Appendix D). Thereby, the domain of $\mathcal{A}$ is given by

$$D(\mathcal{A}) = \left\{ \begin{bmatrix} h_l \\ h_d \end{bmatrix} \in \mathbb{C}^r \oplus W_{1,2}([-T,0];\mathbb{C}^r) \;\middle|\; h_d(0) = h_l \right\}, \tag{2.84}$$

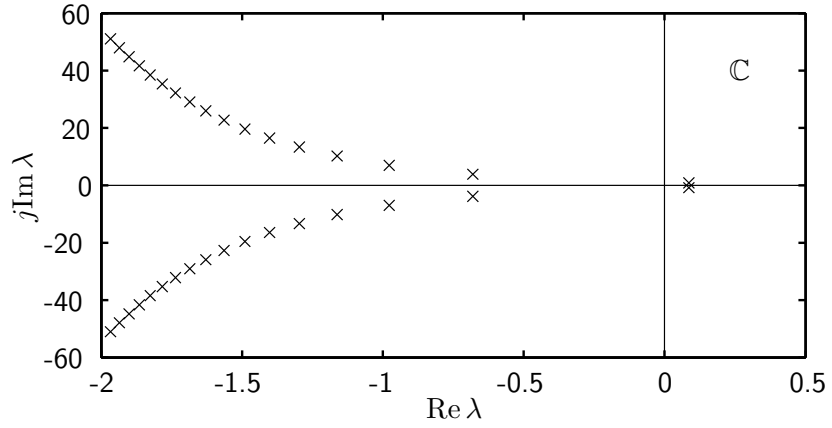

Figure 8 – Eigenvalue distribution of the time-delay system (2.78)–(2.79) with $A_l = A_d = -1$, $B_l = C_l = 1$, and $T = 2$.

(for the definition of $W_{1,2}([-T, 0]; \mathbb{C}^r)$ see Appendix D) which reflects the boundary condition (2.76). Now, it will be checked if this system satisfies the Assumptions 2.1-2 and 2.1-9. When the matrices B_l and C_l are decomposed according to

$$B_l = \begin{bmatrix} b_{l,1} & \cdots & b_{l,p} \end{bmatrix}, \qquad C_l = \begin{bmatrix} c_{l,1}^T \\ \vdots \\ c_{l,m}^T \end{bmatrix} \tag{2.85}$$

and the inner product on X is introduced as

$$\left\langle \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \right\rangle_X := g_1^T \overline{h_1} + \int_{-T}^{0} g_2^T(\tau) \overline{h_2(\tau)} \mathrm{d}\tau, \tag{2.86}$$

then $\mathcal{B}$ and $\mathcal{C}$ according to (2.81)–(2.82) can be written in the form (2.14)–(2.15) with

$$b_i = \begin{bmatrix} b_{l,i} \\ 0 \end{bmatrix}, \quad i = 1, 2, \ldots, p, \qquad c_i = \begin{bmatrix} \overline{c_{l,i}} \\ 0 \end{bmatrix}, \quad i = 1, 2, \ldots, m. \tag{2.87}$$

Thus, Assumption 2.1-2 holds. With the inner product (2.86) the system (2.78)–(2.79) belongs to a class of time-delay systems for which it is shown in [46, Thm. 2.4.4 and Thm. 5.1.7] that the Items 1 and 2 of Assumption 2.1-9 are satisfied. Further, it is stated in [46, Thm. 2.4.6] that the spectrum of $\mathcal{A}$ is discrete, consisting of eigenvalues with finite multiplicities so that the Items 3a, 3b, 3c, and 3d are satisfied. Thus, both Assumptions 2.1-2 and 2.1-9 are verified so that the system (2.78)–(2.79) belongs to the class of systems within the scope of this thesis. The eigenvalue distribution for the simple case $A_l = A_d = -1$, $B_l = C_l = 1$, and $T = 2$ is depicted in Figure 8. ◀

2.2 Modal system approximation

The basic idea behind most approximation schemes is to reduce the dimension of the infinite-dimensional state by projecting it onto a finite-dimensional *linear subspace*[18] X_n of the state space X. Thus, one considers the projected state $x_n(t) = \mathcal{P}x(t)$, where $\mathcal{P}$ denotes the *projection (operator)*[19] onto the subspace X_n. The dynamics of x_n can be described by a finite-dimensional model which approximates the infinite-dimensional system. Besides the question on which subspace the state x is projected, it is an important issue to decide which part of the state space X shall be truncated by the projection because the part x_r of the state x within this subspace X_r will be neglected by the approximation. In terms of the projection this means that its *range space*[20] $\operatorname{ran}\mathcal{P} = X_n$ and its *null space*[21] $\operatorname{nul}\mathcal{P} = X_r$ have to be chosen suitably. This freedom in X_n and X_r offers the possibility to determine approximations with certain properties. A general setting that provides this kind of degrees of freedom is the *Petrov-Galerkin approximation* (see [82]). The model reduction approach considered throughout this thesis is the classical *modal approximation* for which X_n and X_r are *modal subspaces*. These are subspaces of X that are spanned by eigenvectors of $\mathcal{A}$ or are $\mathcal{A}$-*invariant*[22] more generally. In this case all the eigenvalues of the approximation coincide with some of the infinite-dimensional system. For this reason approximations of stable systems are again stable, which is not guaranteed in general for other types of approximations. Moreover, only for modal subspaces X_n and X_r it is possible to describe the dynamics of the projected state $x_n(t) = \mathcal{P}x(t)$ by means of a finite-dimensional model exactly.

For what follows, the eigenvalues λ_i of $\mathcal{A}$ are divided into the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ that will be incorporated in the modal approximation and the remaining eigenvalues $\lambda_{n+1}, \lambda_{n+2}, \ldots$ that will be neglected in the approximation. In the following subsection the modal approximation scheme is explained for system operators whose eigenvectors

[18] A subset $\tilde{H}$ of a linear space H over $\mathbb{C}$ is called a *linear subspace* of H if $\alpha g + \beta h \in \tilde{H}$ holds for all $g, h \in \tilde{H}$ and $\alpha, \beta \in \mathbb{C}$. Then, $\tilde{H}$ itself is a linear space over $\mathbb{C}$.

[19] A linear operator $\mathcal{P} : D(\mathcal{P}) = X \mapsto X$ is called a *projection* if $\mathcal{P}^2 = \mathcal{P}$ is satisfied. It is said to project *onto* X_n *along* X_r if $\operatorname{ran}\mathcal{P} = X_n$ and $\operatorname{nul}\mathcal{P} = X_r$.

[20] The *range space* $\operatorname{ran}\mathcal{M}$ of an operator $\mathcal{M} : D(\mathcal{M}) \subseteq H_1 \mapsto H_2$ is the set of images of $D(\mathcal{M})$ under $\mathcal{M}$, *i.e.*, $\operatorname{ran}\mathcal{M} := \{h \in X \,|\, \exists\, g \in D(\mathcal{M}) \,:\, h = \mathcal{M}g\}$.

[21] The *null space* $\operatorname{nul}\mathcal{M}$ of an operator $\mathcal{M} : D(\mathcal{M}) \subseteq H_1 \mapsto H_2$ is the subset of $D(\mathcal{M})$ for whose elements their images vanish, *i.e.*, $\operatorname{nul}\mathcal{M} := \{h \in D(\mathcal{M}) \,|\, \mathcal{M}h = 0\}$.

[22] A linear subspace $\tilde{X}$ of X is said to be $\mathcal{A}$-*invariant* if $\mathcal{A}h \in \tilde{X}$ holds for all $h \in \tilde{X} \cap D(\mathcal{A})$.

are a Riesz basis. Afterwards in Subsection 2.2.2, the approach is extended to system operators whose eigenvectors are not a basis for X.

2.2.1 Systems with eigenvector Riesz basis

It has been said already before that the eigenvectors ϕ_i of the system operator build a Riesz basis—after suitable scaling—in many applications, which is the case especially for Riesz-spectral operators. This situation will be assumed in this subsection, in addition to Assumption 2.1-9. The projection $\mathcal{P}$ that maps $x(t)$ onto a subspace X_n along a subspace X_r apparently must have the property

$$\mathcal{P}x(t) = x_n(t) \tag{2.88}$$

for all

$$x(t) = x_n(t) + x_r(t), \qquad x_n(t) \in X_n,\ x_r(t) \in X_r, \tag{2.89}$$

where X_n and X_r are the modal subspaces

$$X_n := \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\}, \qquad X_r := \overline{\operatorname{span}}_{i>n}\{\phi_i\} \tag{2.90}$$

that correspond to the dominant eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ and the remaining eigenvalues, respectively. Note, that the decomposition (2.89) is unique for every $x(t) \in X$ because it was assumed that the eigenvectors ϕ_i, $i \in \mathbb{N}$, form a Riesz basis and due to $X_n \cap X_r = \{0\}$ (see [106, Thm. 4.11.3]). In other words, X has the *internal direct sum*[23] $X = X_n + X_r$.

An explicit expression for $\mathcal{P}$ is given by

$$\mathcal{P}h = \sum_{i=1}^{n} \langle h, \psi_i \rangle_X \phi_i, \qquad \forall\, h \in X, \tag{2.91}$$

where $\{\psi_i, i \in \mathbb{N}\}$ is the biorthonormal sequence associated with $\{\phi_i, i \in \mathbb{N}\}$ (see Subsection 2.1.2). In order to justify this expression (2.88) will be verified. Writing $x(t) \in X$ as

$$x(t) = \sum_{i=1}^{\infty} \langle x(t), \psi_i \rangle_X \phi_i \tag{2.92}$$

[23] Let H be a Hilbert space and H_1, H_2 linear subspaces of it. Then, H is said to be the *internal direct sum* of H_1 and H_2, denoted $H = H_1 + H_2$, if every $h \in H$ has a unique decomposition $h = h_1 + h_2$ with $h_1 \in H_1$ and $h_2 \in H_2$.

(see (2.22)) immediately leads to the decomposition (2.89) with

$$x_n(t) = \sum_{i=1}^{n} \langle x(t), \psi_i \rangle_X \phi_i, \qquad x_r(t) = \sum_{i=n+1}^{\infty} \langle x(t), \psi_i \rangle_X \phi_i, \tag{2.93}$$

and application of (2.91) and use of $\langle \phi_i, \psi_j \rangle_X = \delta_{ij}$ (see (2.21)) confirms (2.88). Now, the projection $\mathcal{P}$ is used for determining the modal approximation model for $x_n(t) = \mathcal{P}x(t)$. From (2.3) and $x(t) = x_n(t) + x_r(t)$ it follows

$$\mathcal{P}\dot{x}(t) = \mathcal{P}\mathcal{A}x(t) + \mathcal{P}\mathcal{B}u(t) = \mathcal{P}\mathcal{A}x_n(t) + \mathcal{P}\mathcal{A}x_r(t) + \mathcal{P}\mathcal{B}u(t). \tag{2.94}$$

Since ϕ_i, $i > n$, are eigenvectors of $\mathcal{A}$ it is straightforward to verify that X_r is $\mathcal{A}$-invariant, *i.e.*, $\mathcal{A}h \in X_r$ for all $h \in X_r \cap D(\mathcal{A})$. In view of $X_r = \operatorname{nul} \mathcal{P}$ (see (2.88)–(2.89)) it follows therefore

$$\mathcal{P}\mathcal{A}x_r(t) = 0, \qquad t \geq 0. \tag{2.95}$$

Similar, X_n is $\mathcal{A}$-invariant, *i.e.*, $\mathcal{A}h \in X_n$ for all $h \in X_n \cap D(\mathcal{A})$, which by aid of (2.88)–(2.89) leads to

$$\mathcal{P}\mathcal{A}x_n(t) = \mathcal{A}x_n(t) = \mathcal{A}\mathcal{P}x(t), \qquad t \geq 0. \tag{2.96}$$

Thus, inserting (2.95)–(2.96) into (2.94) yields

$$\mathcal{P}\dot{x}(t) = \mathcal{A}\mathcal{P}x(t) + \mathcal{P}\mathcal{B}u(t), \qquad t > 0 \tag{2.97}$$

so that the dynamics of $x_n(t) = \mathcal{P}x(t)$ is given by

$$\dot{x}_n(t) = \mathcal{A}x_n(t) + \mathcal{P}\mathcal{B}u(t), \qquad t > 0, \qquad x_n(0) = x_{n,0} = \mathcal{P}x_0. \tag{2.98}$$

Note, that this model describes the dynamics of the projected state $x_n(t)$ exactly which is not possible to achieve if X_n and X_r are non-modal. For the purpose of the controller design it is desirable to have a lumped-parameter model Σ_n on the state space $\mathbb{C}^n$ in the usual form

$$\Sigma_n: \quad \dot{\xi}_n(t) = A_n\xi_n(t) + B_n u(t), \qquad t > 0, \quad \xi_n(0) = \xi_{n,0} \in \mathbb{C}^n \tag{2.99}$$

$$y_n(t) = C_n\xi_n(t), \qquad t \geq 0, \tag{2.100}$$

where A_n, B_n, and C_n are matrices of appropriate dimensions. In order to convert (2.98) to the state equation (2.99) note that (2.88) and (2.91) yield

$$x_n(t) = \mathcal{P}x(t) = \sum_{i=1}^{n} x_i^*(t)\phi_i, \tag{2.101}$$

where

$$x_i^*(t) = \langle x(t), \psi_i \rangle_X, \qquad i \in \mathbb{N} \tag{2.102}$$

are called *modal states*. Under use of (2.17) one obtains

$$\mathcal{A}x_n(t) = \sum_{i=1}^{n} \mathcal{A}x_i^*(t)\phi_i = \sum_{i=1}^{n} \lambda_i x_i^*(t)\phi_i \tag{2.103}$$

from (2.101). In addition, (2.14) and (2.91) yield

$$\mathcal{P}\mathcal{B}u(t) = \sum_{i=1}^{n} \begin{bmatrix} \langle b_1, \psi_i \rangle_X & \cdots & \langle b_p, \psi_i \rangle_X \end{bmatrix} \phi_i \, u(t), \tag{2.104}$$

and (2.101) gives

$$x_n(0) = \sum_{i=1}^{n} x_i^*(0)\phi_i. \tag{2.105}$$

Thus, insertion of (2.101)–(2.105) into (2.98), combined with equating coefficients with respect to ϕ_i, leads to

$$\dot{x}_i^*(t) = \lambda_i x_i^*(t) + \begin{bmatrix} \langle b_1, \psi_i \rangle_X & \cdots & \langle b_p, \psi_i \rangle_X \end{bmatrix} u(t), \qquad x_i^*(0) = \langle x(0), \psi_i \rangle_X \tag{2.106}$$

for $i = 1, 2, \ldots, n$. When these equations are written in vector form with $\xi_n(t) = [x_1^*(t) \;\; x_2^*(t) \;\; \cdots \;\; x_n^*(t)]^T$, and the approximative output

$$y_n(t) := \mathcal{C}\mathcal{P}x(t) = \mathcal{C}x_n(t) = \sum_{i=1}^{n} \mathcal{C}\phi_i x_i^*(t) = \begin{bmatrix} \mathcal{C}\phi_1 & \cdots & \mathcal{C}\phi_n \end{bmatrix} \xi_n(t) \tag{2.107}$$

is introduced, the approximation model Σ_n is given by (2.99)–(2.100) and

$$A_n = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \tag{2.108}$$

$$B_n = \begin{bmatrix} \langle b_1, \psi_1 \rangle_X & \cdots & \langle b_p, \psi_1 \rangle_X \\ \vdots & & \vdots \\ \langle b_1, \psi_n \rangle_X & \cdots & \langle b_p, \psi_n \rangle_X \end{bmatrix} \tag{2.109}$$

$$C_n = \begin{bmatrix} \mathcal{C}\phi_1 & \cdots & \mathcal{C}\phi_n \end{bmatrix} \tag{2.110}$$

$$\xi_{n,0} = \begin{bmatrix} \langle x_0, \psi_1 \rangle_X \\ \vdots \\ \langle x_0, \psi_n \rangle_X \end{bmatrix}. \tag{2.111}$$

Next, the dynamics of the *residual state*

$$x_r(t) = (I - \mathcal{P})x(t) = x(t) - x_n(t) \tag{2.112}$$

is considered. By an analog reasoning as in (2.94)–(2.98) and (2.107) this dynamics as well as their contribution y_r to the output y can be described by

$$\dot{x}_r(t) = \mathcal{A}x_r(t) + (I - \mathcal{P})\mathcal{B}u(t), \qquad x_r(0) = (I - \mathcal{P})x_0 \in X_r \tag{2.113}$$

$$y_r(t) = \mathcal{C}x_r(t). \tag{2.114}$$

Since $x_r(t)$ belongs to X_r for all $t \geq 0$ the operators $\mathcal{A}$ and $\mathcal{C}$ in this model can be replaced by the *restrictions*[24]

$$\mathcal{A}_r := \mathcal{A}\big|_{D(\mathcal{A})\cap X_r}, \qquad \mathcal{C}_r := \mathcal{C}\big|_{X_r}. \tag{2.115}$$

Thus, by introducing the new input operator

$$\mathcal{B}_r := (I - \mathcal{P})\mathcal{B} \tag{2.116}$$

the residual dynamics (2.113)–(2.114) become

$$\Sigma_r: \quad \dot{x}_r(t) = \mathcal{A}_r x_r(t) + \mathcal{B}_r u(t), \qquad t > 0, \quad x_r(0) = (I - \mathcal{P})x_0 \in X_r \tag{2.117}$$

$$y_r(t) = \mathcal{C}_r x_r(t), \qquad t \geq 0 \tag{2.118}$$

on X_r which is a Hilbert space with the inner product $\langle g, h\rangle_{X_r} = \langle g, h\rangle_X,\ \forall\, g, h \in X_r$. $\mathcal{A}_r$ inherits from $\mathcal{A}$ the property to generate a C_0-semigroup on X_r (see [46, Lem. 2.5.3, Lem. 2.5.8]), and $\mathcal{B}_r$ is obviously bounded as it is $\mathcal{B}$ so that the abstract initial value problem (2.117) is well-posed. Note, that (2.115) implies

$$\sigma(\mathcal{A}_r) = \sigma(\mathcal{A}) \setminus \sigma(A_n) \tag{2.119}$$

by aid of (2.90) which enables to analyze the stability of the residual dynamics. For doing so, it is important that the SDGA holds for $\mathcal{A}_r$ because its eigenvectors constitute a Riesz basis for X_r (see (2.90) and Proposition 2.1-5). In view of (2.107), (2.112), and (2.114) the output $y(t) = \mathcal{C}x(t) = \mathcal{C}\mathcal{P}x(t) + \mathcal{C}(I - \mathcal{P})x(t)$ has the representation

$$y(t) = y_n(t) + y_r(t), \qquad t \geq 0, \tag{2.120}$$

which shows that the approximation model Σ_n and the residual dynamics Σ_r are complementary with respect to their outputs. Since there is no coupling between the states of the systems Σ_n and Σ_r these two systems can be regarded as a parallel decomposition of Σ as depicted in Figure 9.

[24] An operator $\mathcal{M}_2$ is said to be a *restriction* of an operator $\mathcal{M}_1$, which is written as $\mathcal{M}_2 \subset \mathcal{M}_1$, if $D(\mathcal{M}_2) \subset D(\mathcal{M}_1)$ and $\mathcal{M}_1 h = \mathcal{M}_2 h,\ \forall\, h \in D(\mathcal{M}_2)$. $\mathcal{M}_2 = \mathcal{M}_1|_{H_2}$ denotes the restriction of $\mathcal{M}_1$ with $D(\mathcal{M}_2) = H_2$.

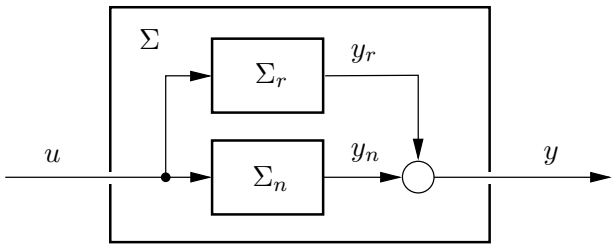

Figure 9 – Parallel decomposition of the infinite-dimensional system Σ into the modal approximation model Σ_n and the residual dynamics Σ_r.

It is interesting to note that projecting the state x onto X_n leads to minimizing the approximation error in the following sense. Introducing the *bijection*[25] $\mathcal{F} : l_2 \mapsto X$, defined by

$$\mathcal{F}h := \sum_{i=1}^{\infty} h_i \phi_i, \qquad \forall\, h \in l_2 \tag{2.121}$$

(for the definition of l_2 see Appendix D) and the coefficient sequences $x^* := (x_1^*, x_2^*, \ldots)$, $x_n^* := (x_1^*, x_2^*, \ldots, x_n^*, 0, 0, \ldots)$, and $x_r^* := (0, \ldots, 0, x_{n+1}^*, x_{n+2}^*, \ldots)$ one has

$$x(t) = \mathcal{F}x^*, \qquad x_n(t) = \mathcal{F}x_n^*, \qquad x_r(t) = \mathcal{F}x_r^* \tag{2.122}$$

due to the unique expansions

$$x(t) = \sum_{i=1}^{\infty} x_i^*(t)\phi_i, \quad x_n(t) = \sum_{i=1}^{n} x_i^*(t)\phi_i, \quad x_r(t) = \sum_{i=n+1}^{\infty} x_i^*(t)\phi_i \tag{2.123}$$

(compare to (2.92)–(2.93) and (2.102)). In the same way as x_n is the projection of x, also x_n^* can be regarded as a projection of x^*, *i.e.*, $x_n^*(t) = \tilde{\mathcal{P}}x^*(t)$. The related projection operator $\tilde{\mathcal{P}} = \mathcal{F}^{-1}\mathcal{P}\mathcal{F}$ is orthogonal because $\langle x_n^*(t), x_r^*(t)\rangle_{l_2} = 0$, $\forall\, t \geq 0$, is obviously satisfied. It is a well-known property of orthogonal projections that the resulting error is minimized (see [106, Thm. 5.16.5]), which means that $x_r^*(t) = x^*(t) - x_n^*(t)$ satisfies

$$\|x_r^*(t)\|_{l_2} \leq \|x^*(t) - \hat{x}_n^*(t)\|_{l_2}, \qquad \forall\, \hat{x}_n^*(t) \in X_n,\ \forall\, t \geq 0. \tag{2.124}$$

In the following example a modal approximation of the one-dimensional heat conductor introduced in Example 2.1-1 is determined.

[25] A linear operator $\mathcal{M} : H_1 \mapsto H_2$ is called a *bijection* if $\mathcal{M}$ is defined and bounded on H_1, and $\mathcal{M}^{-1}$ exists and is defined as a bounded operator on whole H_2.

Example 2.2-1 (1-D heat conductor, continued)

The eigenvalues λ_i and the corresponding eigenvectors ϕ_i of the system operator (2.10) of the heat conductor in Example 2.1-1 can be found as

$$\lambda_i = -i^2\pi^2\frac{\mu}{c\,\rho\,\ell^2}, \qquad i \in \mathbb{N} \tag{2.125}$$

$$\phi_i(z) = \sqrt{\frac{2}{\ell}}\sin\left(\frac{i\pi}{\ell}z\right), \qquad i \in \mathbb{N} \tag{2.126}$$

which form an orthonormal basis for X. Therefore, the biorthonormal sequence $\{\psi_i, i \in \mathbb{N}\}$ corresponding to $\{\phi_i, i \in \mathbb{N}\}$ is simply given by $\psi_i = \phi_i, i \in \mathbb{N}$ (see Footnote 6). Thus, the state $x(t)$ has the series expansion

$$x(t) = \sum_{i=1}^{\infty} x_i^*(t)\,\phi_i, \qquad x_i^*(t) = \langle x(t), \psi_i\rangle_X \tag{2.127}$$

(see (2.22) and (2.102)), where the modal states x_i^* satisfy $x_i^*(t) \to 0$ for $i \to \infty$, $\forall\, t \geq 0$, because the infinite sum would otherwise not converge. It is therefore plausible that the approximation error can be reduced arbitrarily by increasing the approximation order in view of (2.101). Taking $\psi_i = \phi_i$ into account, the approximation (2.108)–(2.111) under use of (2.9), (2.11)–(2.12), and (2.125) becomes

$$A_n = -\frac{\pi^2\,\mu}{c\,\rho\,\ell^2}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 4 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & n^2 \end{bmatrix} \tag{2.128}$$

$$B_n = \sqrt{\frac{2}{\ell}}\frac{1}{c\,\rho}\begin{bmatrix} \int_0^\ell b(z)\sin(\frac{\pi}{\ell}z)\mathrm{d}z \\ \int_0^\ell b(z)\sin(2\frac{\pi}{\ell}z)\mathrm{d}z \\ \vdots \\ \int_0^\ell b(z)\sin(n\frac{\pi}{\ell}z)\mathrm{d}z \end{bmatrix} \tag{2.129}$$

$$C_n = \sqrt{\frac{2}{\ell}}\begin{bmatrix} \int_0^\ell \sin(\frac{\pi}{\ell}z)\overline{c(z)}\mathrm{d}z & \int_0^\ell \sin(2\frac{\pi}{\ell}z)\overline{c(z)}\mathrm{d}z & \cdots & \int_0^\ell \sin(n\frac{\pi}{\ell}z)\overline{c(z)}\mathrm{d}z \end{bmatrix} \tag{2.130}$$

and

$$x_{n,0} = \sqrt{\frac{2}{\ell}}\begin{bmatrix} \int_0^\ell x_0(z)\sin(\frac{\pi}{\ell}z)\mathrm{d}z \\ \int_0^\ell x_0(z)\sin(2\frac{\pi}{\ell}z)\mathrm{d}z \\ \vdots \\ \int_0^\ell x_0(z)\sin(n\frac{\pi}{\ell}z)\mathrm{d}z \end{bmatrix}. \tag{2.131}$$

It is important to note that, depending on $b(z)$ and $c(z)$, the pair (A_n, B_n) may be not controllable and the pair (C_n, A_n) may be not observable. For example, for

$$b(z) = \mathbf{1}_{[\frac{1}{2}-\varepsilon,\frac{1}{2}+\varepsilon]}(z), \qquad c(z) = \mathbf{1}_{[\frac{1}{4}-\varepsilon,\frac{1}{4}+\varepsilon]}(z) \tag{2.132}$$

with $\varepsilon \leq 1/4$ all rows of B_n with index $i = 2, 4, 6, \ldots$ and all columns of C_n with index $i = 4, 8, 12, \ldots$ vanish leading to the mentioned defect of controllability and observability, respectively. Since these do not contribute to the transfer behavior of the system they need not be contained in the approximation. For the specific case $\ell = \mu = c = \rho = 1$ and $b(z)$, $c(z)$ according to (2.132) with $\varepsilon = 0.01$ the step responses of the approximations with orders $n = 2, 3, 4, 5, 6$ are depicted in Figure 10, wherein only the dominant modes that are relevant for the transfer behavior are contained in the approximations. The figure shows that the output trajectories converge quite fast for increasing n. ◀

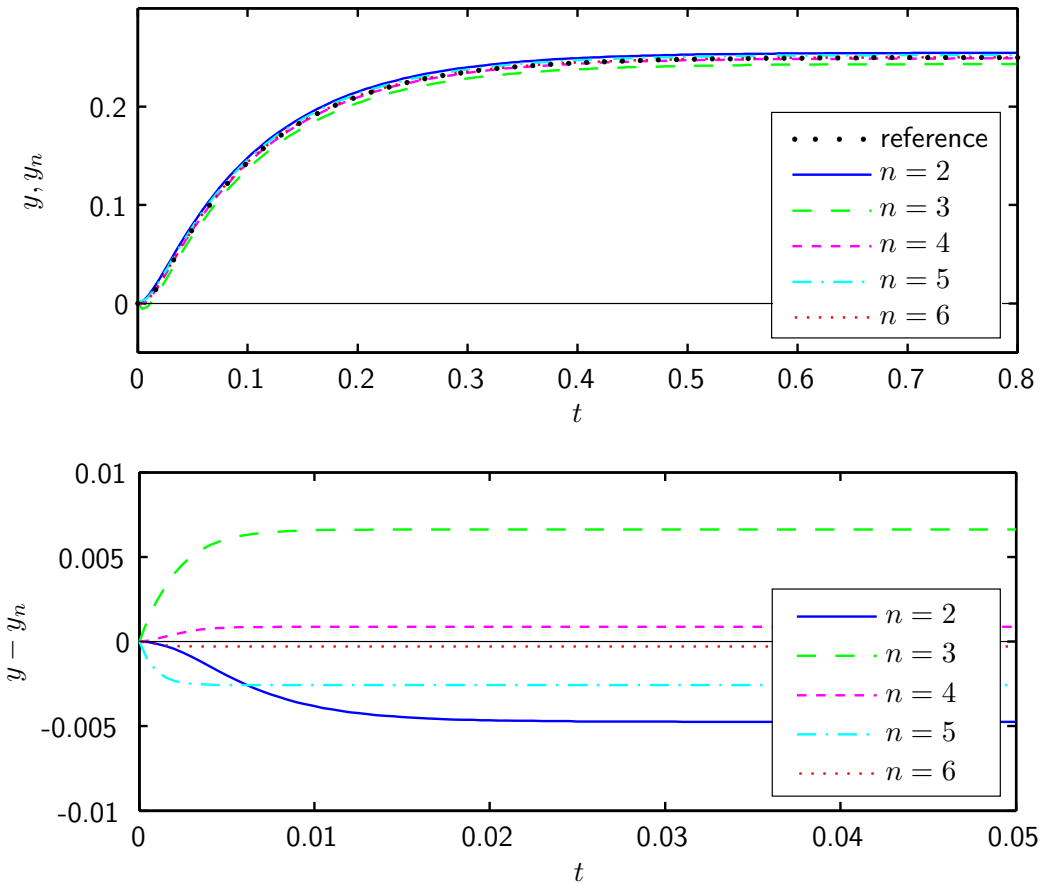

Figure 10 – Step responses y_n of modal approximations with different orders n for the heat conductor as well as the errors $y - y_n$ w.r.t. a reference model with step response y that is a 60-th-order modal approximation.

2.2.2 Systems without eigenvector Riesz basis

In some applications, for instance for time-delay systems, the assumption used in the previous subsection that the eigenvectors of $\mathcal{A}$ form a Riesz basis is violated. This condition is now relaxed by considering instead the following weaker assumption. In this way it is possible to compute modal approximations for systems with eigenvectors ϕ_i that do not span the whole state space X.

Assumption 2.2-2

It is assumed that there exist n linearly independent eigenvectors $\phi_1, \phi_2, \ldots, \phi_n$ of $\mathcal{A}$ that correspond to $\lambda_1, \lambda_2, \ldots, \lambda_n$, where these eigenvalues have finite *algebraic multiplicities*[26] that coincide with their *geometric multiplicities*[27]. In addition, there shall exist n eigenvectors $\tilde{\psi}_1, \tilde{\psi}_2, \ldots, \tilde{\psi}_n$ of an algebraic adjoint $\tilde{\mathcal{A}}^*$ that correspond to $\overline{\lambda_1}, \overline{\lambda_2}, \ldots, \overline{\lambda_n}$. Without loss of generality these are assumed biorthonormalized, *i.e.*, $\langle \phi_i, \tilde{\psi}_j \rangle_X = \delta_{ij}$, $i, j = 1, 2, \ldots, n$. Finally, it is assumed that a rectifiable, closed, simple curve $\Gamma \subset \mathbb{C}$ can be drawn such that it contains $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ in its interior and $\sigma(\mathcal{A}) \setminus \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ in its exterior. ◀

While an *algebraic adjoint* $\tilde{\mathcal{A}}^*$ of $\mathcal{A}$ is defined as an operator $\tilde{\mathcal{A}}^* : D(\tilde{\mathcal{A}}^*) \subset X \mapsto X$ such that

$$\langle \mathcal{A}g, h \rangle_X = \langle g, \tilde{\mathcal{A}}^* h \rangle_X, \qquad \forall\, g \in D(\mathcal{A}),\ \forall\, h \in D(\tilde{\mathcal{A}}^*) \tag{2.133}$$

holds, the *adjoint* $\mathcal{A}^*$ of $\mathcal{A}$ has to fulfill the additional condition that its domain is maximal (see Appendix C). However, since Assumption 2.2-2 involves only an algebraic adjoint the maximal domain needs not to be determined for checking the assumption. It should be remarked that the requirement $\langle \phi_i, \tilde{\psi}_j \rangle_X = \delta_{ij}$ can be achieved always under the mentioned assumptions by suitably scaling the eigenvectors of $\tilde{\mathcal{A}}^*$.

In the previous subsection the property of X_n and X_r to be $\mathcal{A}$-invariant was essential

[26] Let $\mathcal{M} : D(\mathcal{M}) \subseteq H \mapsto H$ denote a linear operator. The *algebraic multiplicity* $\alpha(\lambda)$ of an eigenvalue λ of $\mathcal{M}$ is given by $\alpha(\lambda) = \sum_{k=1}^{\infty} \dim \operatorname{nul} (\lambda I - \mathcal{M})^k$.

[27] Let $\mathcal{M} : D(\mathcal{M}) \subseteq H \mapsto H$ denote a linear operator. The *geometric multiplicity* $\gamma(\lambda)$ of an eigenvalue λ of $\mathcal{M}$ is given by $\gamma(\lambda) = \dim \operatorname{nul} (\lambda I - \mathcal{M})$.

because it implied

$$\dot{x}_n(t) = \mathcal{A}x_n(t) + \mathcal{P}\mathcal{B}u(t) \tag{2.134}$$

$$\dot{x}_r(t) = \mathcal{A}x_r(t) + (I - \mathcal{P})\mathcal{B}u(t) \tag{2.135}$$

(see (2.98) and (2.113)). These equations have the important meaning that the approximation and the residual dynamics are decoupled subsystems. However, the decomposition $X = X_n + X_r$ with X_n and X_r defined in (2.90) is no longer possible if the eigenvectors ϕ_i do not span the whole state space X. Therefore, the question arises in which way X_n, X_r, and the corresponding projection $\mathcal{P}$ have to be redefined such that (2.134)–(2.135) hold. The answer is given in the following statement. Additionally, these subspaces must be $\mathcal{S}$-*invariant*[28] since this property is needed for assuring that $\mathcal{A}_r = \mathcal{A}|_{D(\mathcal{A})\cap X_r}$ is the infinitesimal generator of a C_0-semigroup.

Proposition 2.2-3

Let $X_n := \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\}$ *and* $X_r := \operatorname{span}\{\tilde{\psi}_1, \tilde{\psi}_2, \ldots, \tilde{\psi}_n\}^\perp$, *where* $(\cdot)^\perp$ *denotes the* orthogonal complement[29] *of a linear subspace. Then, the following holds:*

1. $X_n + X_r = X$.
2. *The operator*

$$\mathcal{P}h = \sum_{i=1}^{n} \langle h, \tilde{\psi}_i \rangle_X \phi_i, \qquad \forall\, h \in X \tag{2.136}$$

   is the projection onto X_n *along* X_r.
3. X_n *and* X_r *are* $\mathcal{A}$*-invariant and* $\mathcal{S}$*-invariant.*
4. *(2.134)–(2.135) hold for the redefined projection* $\mathcal{P}$ *according to (2.136).*

For the proof see Appendix A.2. By using this redefined projection $\mathcal{P}$ and the bijection $\mathcal{F}$, redefined by $\mathcal{F} : \mathbb{C}^n \mapsto X_n$ and

$$\mathcal{F}h := \sum_{i=1}^{n} h_i \phi_i, \qquad \forall\, h \in \mathbb{C}^n, \tag{2.137}$$

the approximation Σ_n in (2.99)–(2.100) can be derived in the same way as for systems

[28] Let $\mathcal{S}(t)$ denote a C_0-semigroup on X. A linear subspace $\tilde{X}$ of X is said to be $\mathcal{S}$-*invariant* if $\mathcal{S}(t)h \in \tilde{X}$ holds for all $h \in \tilde{X}$ and all $t \geq 0$.

[29] Let $\tilde{X}$ be a linear subspace of X. Then, the *orthogonal complement* $\tilde{X}^\perp$ is the linear subspace of X defined by $\tilde{X}^\perp := \{g \in X \mid \langle g, h \rangle_X = 0,\ \forall\, h \in \tilde{X}\}$.

with eigenvector basis yielding

$$A_n = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \tag{2.138}$$

$$B_n = \begin{bmatrix} \langle b_1, \tilde{\psi}_1 \rangle_X & \cdots & \langle b_p, \tilde{\psi}_1 \rangle_X \\ \vdots & & \vdots \\ \langle b_1, \tilde{\psi}_n \rangle_X & \cdots & \langle b_p, \tilde{\psi}_n \rangle_X \end{bmatrix} \tag{2.139}$$

$$C_n = \begin{bmatrix} \mathcal{C}\phi_1 & \cdots & \mathcal{C}\phi_n \end{bmatrix} \tag{2.140}$$

$$x_{n,0} = \begin{bmatrix} \langle x_0, \tilde{\psi}_1 \rangle_X \\ \vdots \\ \langle x_0, \tilde{\psi}_n \rangle_X \end{bmatrix} = \mathcal{F}^{-1}\mathcal{P}x_0. \tag{2.141}$$

Finally, the residual dynamics Σ_r is described by (2.113)–(2.114) with the operators

$$\mathcal{A}_r = \mathcal{A}\big|_{D(\mathcal{A})\cap X_r}, \qquad \mathcal{B}_r = (I - \mathcal{P})\mathcal{B}, \qquad \mathcal{C}_r = \mathcal{C}\big|_{X_r} \tag{2.142}$$

as in the case with eigenvector Riesz basis, at which X_r is now redefined as in Proposition 2.2-3. Again, $\mathcal{A}_r$ is the infinitesimal generator of a C_0-semigroup

$$\mathcal{S}_r(t) = (I - \mathcal{P})\mathcal{S}(t), \qquad \forall\, t \geq 0, \tag{2.143}$$

which follows immediately from [46, Lem. 2.5.3]. Furthermore,

$$\sigma(\mathcal{A}_r) = \sigma(\mathcal{A}) \setminus \sigma(A_n) \tag{2.144}$$

is still valid which now follows from (2.142) and Item 1 of Proposition 2.2-3. As explained in Subsection 2.1.3, the spectrum of $\mathcal{A}_r$ determines the stability margin of the generated C_0-semigroup only if $\mathcal{A}_r$ satisfies the SDGA. This condition holds since $\mathcal{A}$ satisfied the SDGA due to Assumption 2.1-9 and in view of (2.143).

It goes without saying that the discussed approach can be applied also for systems with eigenvector Riesz basis because one can show that then $\tilde{\psi}_i = \psi_i$, $i = 1, 2, \ldots, n$, holds so that the approximations for both approaches are the same. For convenience, the state ξ_n of the approximation Σ_n will be denoted x_n from now on. In the following example a modal approximation for a system without complete eigenvector basis is determined.

Example 2.2-4 (Time-delay system, continued)

The time-delay system of Example 2.1-16 is considered for which the operators

$$\mathcal{A}\begin{bmatrix} h_l \\ h_d \end{bmatrix} = \begin{bmatrix} A_l h_l + A_d h_d(-T) \\ \mathrm{d}_\tau h_d \end{bmatrix}, \qquad \forall \begin{bmatrix} h_l \\ h_d \end{bmatrix} \in D(\mathcal{A}) \tag{2.145}$$

$$D(\mathcal{A}) = \left\{ \begin{bmatrix} h_l \\ h_d \end{bmatrix} \in \mathbb{C}^r \oplus W_{1,2}([-T,0];\mathbb{C}^r) \;\middle|\; h_d(0) = h_l \right\} \tag{2.146}$$

$$\mathcal{B}v = \begin{bmatrix} B_l \\ 0 \end{bmatrix} v, \qquad \forall\, v \in \mathbb{C} \tag{2.147}$$

$$\mathcal{C}h = \begin{bmatrix} C_l & 0 \end{bmatrix} h, \qquad \forall\, h \in X \tag{2.148}$$

of the state space model (2.3)–(2.4) have been found, wherein

$$X := \mathbb{C}^r \oplus L_2([-T,0];\mathbb{C}^r) \tag{2.149}$$

is the state space. For the computation of the approximation according to (2.138)–(2.141) the eigenvectors ϕ_i, $i = 1, 2, \ldots, n$, of $\mathcal{A}$ and the corresponding eigenvectors $\tilde{\psi}_i$ of an algebraic adjoint $\tilde{\mathcal{A}}^*$ must be determined. The eigenvalue-eigenvector equation

$$(\lambda I - \mathcal{A})\begin{bmatrix} \phi_l \\ \phi_d \end{bmatrix} = \begin{bmatrix} (\lambda - A_l)\phi_l - A_d\phi_d(-T) \\ \lambda\phi_d - \mathrm{d}_\tau\phi_d \end{bmatrix} = 0 \tag{2.150}$$

shows that ϕ_d has to satisfy

$$\phi_d(\tau) = \phi_l e^{\lambda\tau} \tag{2.151}$$

(see the second row of (2.150)), where $\phi_d(0) = \phi_l$, following from $D(\mathcal{A})$, has been used. The first row of (2.150), combined with (2.151), yields

$$(\lambda I - A_l)\phi_l - A_d\phi_d(-T) = (\lambda I - A_l - A_d e^{-\lambda T})\phi_l = 0. \tag{2.152}$$

This equation is satisfied if ϕ_l is an eigenvector of the matrix $A_l + A_d e^{-\lambda T}$ for the eigenvalue λ, where it is a necessary condition that the characteristic equation

$$\det(\lambda I - A_l - A_d e^{-\lambda T}) = 0 \tag{2.153}$$

holds, whose solutions are the eigenvalues λ_i, $i \in \mathbb{N}$, of $\mathcal{A}$. Thus,

$$\phi_i(\tau) = \begin{bmatrix} \phi_{l,i} \\ \phi_{l,i} e^{\lambda_i \tau} \end{bmatrix}, \qquad i \in \mathbb{N} \tag{2.154}$$

are eigenvectors of $\mathcal{A}$ with the corresponding eigenvalues λ_i satisfying (2.153) and $\phi_{l,i}$ satisfying (2.152). For the computation of the vectors $\tilde{\psi}_i$ appearing in the approximation an algebraic adjoint $\tilde{\mathcal{A}}^*$ has to be found. A straightforward calculation shows that

$$\tilde{\mathcal{A}}^* \begin{bmatrix} h_l \\ h_d \end{bmatrix} = \begin{bmatrix} \overline{A_d}^T h_l + h_d(0) \\ -\mathrm{d}_\tau h_d \end{bmatrix}, \qquad \forall \begin{bmatrix} h_l \\ h_d \end{bmatrix} \in D(\tilde{\mathcal{A}}^*) \tag{2.155}$$

$$D(\tilde{\mathcal{A}}^*) = \left\{ \begin{bmatrix} h_l \\ h_d \end{bmatrix} \in \mathbb{C}^r \oplus W_{1,2}([-T,0];\mathbb{C}^n) \;\middle|\; h_d(-T) = \overline{A_d}^T h_l \right\} \tag{2.156}$$

satisfies the defining relation $\langle \mathcal{A}g, h\rangle_X = \langle g, \tilde{\mathcal{A}}^* h\rangle_X$, $\forall\, g \in D(\mathcal{A}), \forall\, h \in D(\tilde{\mathcal{A}}^*)$, of an algebraic adjoint. The eigenvectors $\tilde{\psi}_i$ of $\tilde{\mathcal{A}}^*$ that correspond to $\overline{\lambda_i}$ can be determined in an analog way as the vectors ϕ_i above which yields

$$\tilde{\psi}_i(\tau) = \alpha_i \begin{bmatrix} \tilde{\psi}_{l,i} \\ \overline{A_d}^T \tilde{\psi}_{l,i} e^{-\overline{\lambda_i}(T+\tau)} \end{bmatrix}, \tag{2.157}$$

where $\tilde{\psi}_{l,i}$ are eigenvectors of the matrix $\overline{A_l}^T + \overline{A_d}^T e^{-\overline{\lambda_i} T}$ for the eigenvalue $\overline{\lambda_i}$, and $\alpha_i \in \mathbb{C}$ are scaling factors. For

$$\alpha_i = \frac{1}{\phi_{l,i}^T (I + A_d^T T e^{-\lambda_i T}) \overline{\tilde{\psi}_{l,i}}} \tag{2.158}$$

the sequences $\{\phi_i, i \in \mathbb{N}\}$ and $\{\tilde{\psi}_i, i \in \mathbb{N}\}$ are biorthonormal. Inserting ϕ_i and $\tilde{\psi}_i$ according to (2.154) and (2.157)–(2.158), respectively, into (2.138)–(2.140) and using (2.147)–(2.148) and (2.86) finally yields the approximation Σ_n with

$$A_n = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \tag{2.159}$$

$$B_n = \begin{bmatrix} \overline{\alpha_1 \tilde{\psi}_{l,1}}^T \\ \vdots \\ \overline{\alpha_n \tilde{\psi}_{l,n}}^T \end{bmatrix} B_l \tag{2.160}$$

$$C_n = C_l \begin{bmatrix} \phi_{l,1} & \cdots & \phi_{l,n} \end{bmatrix}. \tag{2.161}$$

For the special case $A_l = A_d = -1$, $B_l = C_l = 1$, and $T = 2$ some of the eigenvalues with largest real parts are computed by numerically solving (2.153) and shown in Table 1 and Figure 8. In Figure 11 the Bode plots of a 6-th-order modal approximation is shown. As expected, the transfer behavior of the modal approximation in the frequency regions close to the eigenvalues $\lambda_{\pm i}$, $i = 1, 2, 3$, is in accordance with the behavior of the original system. ◀

Table 1 – Eigenvalues of $\mathcal{A}$ with largest real parts. Due to $\mathrm{Re}\,\lambda_{\pm 1} > 0$ the system is unstable. See also Figure 8.

i	1	2	3	4
$\lambda_{\pm i}$	$0.086 \pm j0.837$	$-0.680 \pm j3.839$	$-0.978 \pm j6.999$	$-1.162 \pm j10.153$

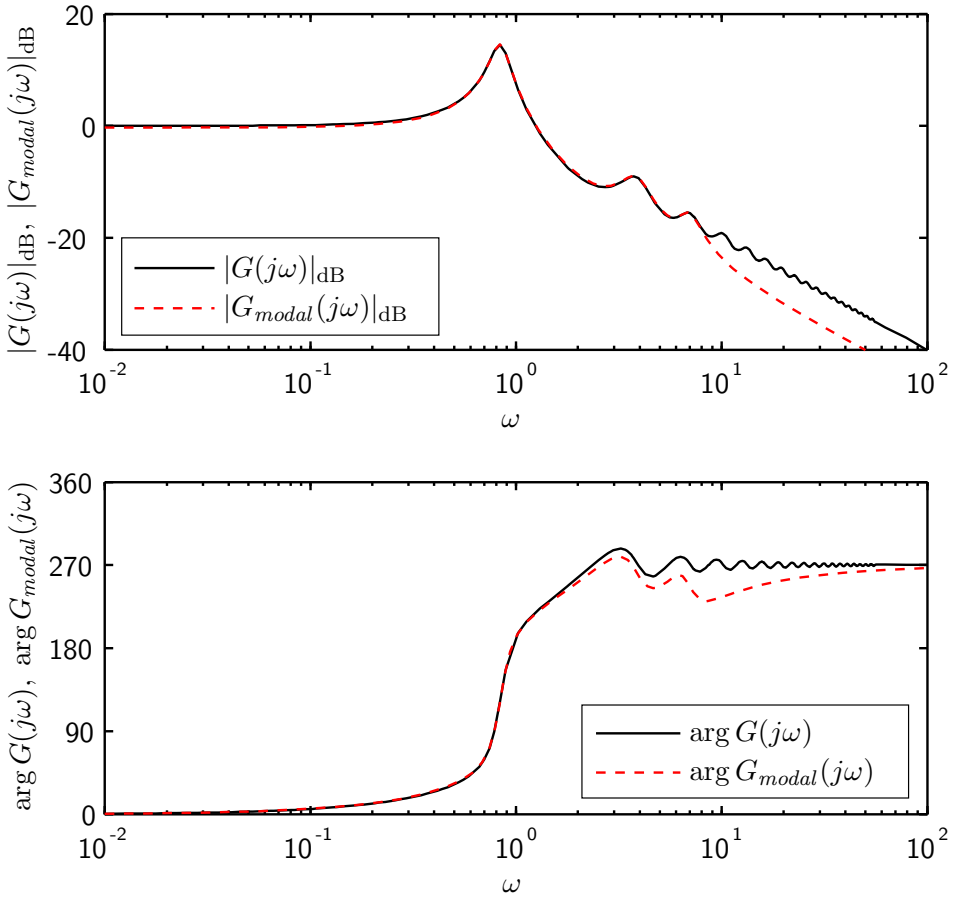

Figure 11 – Bode plots of the infinite-dimensional time-delay system ($|G(j\omega)|_{\mathrm{dB}}$) and the 6-th-order modal approximation ($|G_{modal}(j\omega)|_{\mathrm{dB}}$).

2.3 Observer-based compensator design

In this section the control of infinite-dimensional systems Σ of the form (2.3)–(2.4) by means of state feedback control is considered. The feedback law is implemented by a Luenberger observer, which, combined with the state feedback, forms an observer-based compensator. Thus, the closed-loop has the structure as depicted in Figure 12a. In many applications it is desired not only to assure the stability of the closed-loop system and to assign a certain stability margin but, in addition, to achieve a specified reference tracking performance. For this purpose the *two-degrees-of-freedom control*, as shown in Figure 12b, provides a suitable control scheme. While the tracking behavior w.r.t. the reference input w is adjusted by the feedforward control Σ_{ff}, errors between the desired output trajectory y_d and the actual plant output signal y are regulated by the feedback control Σ_c. It is a benefit of the two-degrees-of-freedom structure that feedforward control and feedback control can be designed independently from each other (see [51]). Since the spillover effect basically endangers the stability of the closed-loop system, this thesis focuses on the stabilization issue, for which purpose the simpler structure in Figure 12a is considered in what follows. Nevertheless, the feedforward control Σ_{ff}, that does not affect the closed-loop stability, can be designed separately (see [51, 114]).

For the design of the observer-based compensator a modal approximation Σ_n is used. Hence, the standard design approaches for finite-dimensional LTI systems can be applied, which is the basic concept of the early-lumping approach. The fundamental problem of doing so is that the observer expects to receive the output y_n of the approximation model. However, since this output is not available, the observer is fed instead with the measurable output y of the infinite-dimensional system Σ. As a con-

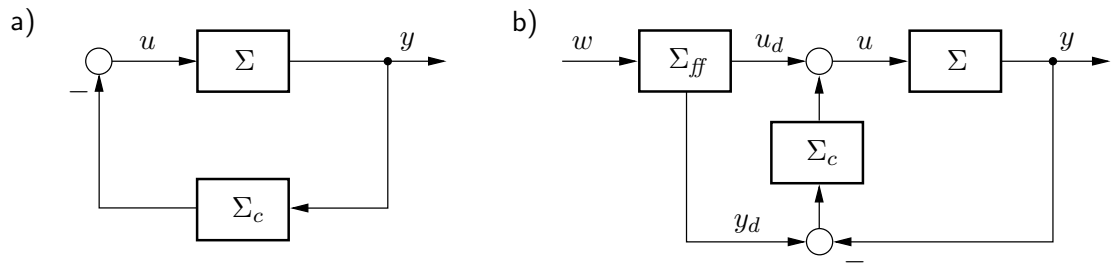

Figure 12 – a) Structure of the control loop with plant Σ and observer-based compensator Σ_c. b) Control loop in two-degrees-of-freedom structure with feedforward control Σ_{ff} and compensator Σ_c, besides plant Σ.

sequence, the observation error is excited by the residual dynamics which in turn affects the closed-loop behavior. Due to this effect, which is introduced in [5] as *observation spillover*, the closed-loop dynamics has to be checked after the compensator has been designed.

In the following subsection the requirements concerning stabilizability and detectability of the infinite-dimensional system are discussed. Afterwards, the design of the compensator on the basis of the closed-loop structure as shown in Figure 12a is considered in Subsection 2.3.2, and the closed-loop dynamics is analyzed in Subsection 2.3.3, where the effect of spillover becomes apparent. Finally, the asymptotic rejection of exogenous disturbances is considered briefly in Subsection 2.3.4.

2.3.1 Stabilizability and detectability

By an appropriately designed compensator it shall be possible to achieve a prescribed *stability margin* $\beta > 0$ for the closed-loop dynamics which is defined as the negative growth bound of the corresponding C_0-semigroup[30]. To this end, the system Σ in (2.3)–(2.4) is assumed to be *exponentially β-stabilizable* in the following. This means that it is possible to apply the state feedback $u(t) = -\mathcal{K}x(t)$ with a bounded linear operator $\mathcal{K} : X \mapsto \mathbb{C}^p$ to (2.3) such that the resulting closed-loop system operator $\mathcal{A} - \mathcal{B}\mathcal{K}$ generates a C_0-semigroup with a growth bound $\omega_0 \leq -\beta$. This requires to shift all eigenvalues that are located in the right half-plane

$$\mathbb{C}^+_{-\beta} := \{s \in \mathbb{C} \,|\, \operatorname{Re} s > -\beta\} \tag{2.162}$$

into the left half-plane

$$\overline{\mathbb{C}}^-_{-\beta} := \{s \in \mathbb{C} \,|\, \operatorname{Re} s \leq -\beta\}. \tag{2.163}$$

The fact that the considered systems Σ have a bounded input operator $\mathcal{B}$ and a finite numbers p of inputs implies that the input operator $\mathcal{B}$ has *finite rank*[31]. This has the consequence that only finitely many eigenvalues of $\mathcal{A}$ can be shifted by the compensator from $\mathbb{C}^+_{-\beta}$ to $\overline{\mathbb{C}}^-_{-\beta}$ (see [46, Thm. 5.2.6]). Therefore, a system Σ can be exponentially β-stabilizable only if $\mathbb{C}^+_{-\beta}$ contains not more than a finite number of spectral points. Of

[30] That means that the *stability margin* β of a C_0-semigroup $\mathcal{S}_{\mathcal{M}}(t)$ on X is the supremum of all real numbers $\tilde{\beta}$ such that a constant C exists with $\|\mathcal{S}(t)x_0\|_X \leq Ce^{-\tilde{\beta}t}\|x_0\|_X$ for all $x_0 \in X$ and all $t \geq 0$.

[31] An operator $\mathcal{M} : H_1 \mapsto H_2$ is said to have a *finite rank* if $\dim \operatorname{ran} \mathcal{M} < \infty$.

course, this requirement excludes systems with an accumulation point *within* the half-plane $\mathbb{C}^+_{-\beta}$ as may be the case, *e.g.*, for the Euler-Bernoulli beam with sufficiently strong Kelvin-Voigt damping, as well as systems whose eigenvalues are entirely located on a vertical line in $\mathbb{C}^+_{-\beta}$, such as the undamped Euler-Bernoulli beam. Since the observer-based compensator will be designed by aid of a modal approximation Σ_n only those eigenvalues, whose corresponding *eigenmodes*[32] are contained in the approximation, can be incorporated in the design. Therefore, the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A_n must comprise all unstable eigenvalues of $\mathcal{A}$ as well as those that are intended to be shifted by means of the control. The remaining eigenvalues $\lambda_{n+1}, \lambda_{n+2}, \ldots$ of the residual dynamics, *i.e.*, the eigenvalues of $\mathcal{A}_r$, which are not contained in the approximation, must have sufficiently negative real parts so that they can be left unchanged. To be more precise, the decomposition

$$\sigma(A_n) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \subset \mathbb{C}^+_{-\beta}, \qquad \sigma(\mathcal{A}_r) \subset \overline{\mathbb{C}}^-_{-\beta} \tag{2.164}$$

of $\sigma(\mathcal{A}) = \sigma(A_n) \cup \sigma(\mathcal{A}_r)$ is assumed (see (2.144)). This assures that the residual dynamics have a stability margin of at least β so that these dynamics need not be changed by the control. This conclusion holds true, however, only if $\mathcal{A}_r$ satisfies the SDGA, which will be assured later by assuming the SDGA for the closed-loop system (see Assumption 2.3-2).

The following considerations require that multiple eigenvalues are repeated accordingly to their multiplicities. That means that the eigenvalues $\{\lambda_{q_1}, \lambda_{q_2}, \ldots, \lambda_{q_r}\} \subseteq \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ that are mutual different, *i.e.*, $\lambda_{q_i} \neq \lambda_{q_j}$ for $i \neq j$, are repeated as often as the algebraic multiplicity ν_i of λ_{q_i} so that

$$\lambda_{q_i} = \lambda_{q_i+1} = \ldots = \lambda_{q_i+\nu_i-1}, \qquad i = 1, 2, \ldots, r \tag{2.165}$$

holds. An easily verifiable condition for Σ to be exponentially β-stabilizable can be obtained from the fact that this property is equivalent in this set up to the controllability of the modal approximation Σ_n (see [46, Thm. 5.2.6]). Using the *Gilbert controllability criterion* it is straightforward to see in view of (2.108)–(2.109) that this in turn is equivalent to

$$\operatorname{rank} \begin{bmatrix} \langle b_1, \psi_{q_i} \rangle_X & \cdots & \langle b_p, \psi_{q_i} \rangle_X \\ \vdots & & \vdots \\ \langle b_1, \psi_{q_i+\nu_i-1} \rangle_X & \cdots & \langle b_p, \psi_{q_i+\nu_i-1} \rangle_X \end{bmatrix} = \nu_i, \qquad \forall\, i = 1, 2, \ldots, r \tag{2.166}$$

[32] By the term *eigenmode* the subsystem of a state space system is meant that corresponds to a certain single eigenvalue of the system operator.

(see [86, Sec. 6.2]) under which condition the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are said to be *modal controllable*. Of course, ψ_{q_i} therein has to be replaced by $\tilde{\psi}_{q_i}$ if the approximation (2.138)–(2.141) is used.

For achieving the stability margin β of the closed-loop system it has to be assumed in addition to the β-stabilizability that the system Σ is *exponentially β-detectable* which means that it is possible to find a bounded linear operator $\mathcal{L} : \mathbb{C}^m \mapsto X$ such that the C_0-semigroup corresponding to $\mathcal{A} - \mathcal{L}\mathcal{C}$ has a growth bound $\omega_0 \leq -\beta$. This is guaranteed if and only if the mentioned approximation is observable (see [46, Thm. 5.2.10]). In view of (2.108) and (2.110) this in turn is equivalent to

$$\operatorname{rank} \begin{bmatrix} \mathcal{C}\phi_{q_i} & \cdots & \mathcal{C}\phi_{q_i+\nu_i-1} \end{bmatrix} = \nu_i, \qquad \forall\, i = 1, 2, \ldots, r \tag{2.167}$$

by the *Gilbert observability criterion* (see [86, Sec. 6.2]). In this case, the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are called *modal observable*. Summing up these considerations, the compensator design will be based on the following assumption.

Assumption 2.3-1 (Exponential β-stabilizability and β-detectability)

For the system (2.3)–(2.4) the following is assumed in order to assure exponential β-stabilizability and exponential β-detectability:

1. There are not more than finitely many eigenvalues located in the half-plane $\mathbb{C}^+_{-\beta} = \{s \in \mathbb{C} \,|\, \operatorname{Re} s > -\beta\}$, where $\beta > 0$ is the desired stability margin of the closed-loop system.
2. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are modal controllable and modal observable, *i.e.*, (2.166)–(2.167) are satisfied. ◀

2.3.2 Design of the observer-based compensator

Now, the *observer-based compensator*

$$\Sigma_c : \; \dot{\hat{x}}_n(t) = (A_n - LC_n)\hat{x}_n(t) + B_n u(t) + Ly(t), \quad t > 0, \; \hat{x}_n(0) = \hat{x}_{n,0} \in \mathbb{C}^n \tag{2.168}$$

$$u(t) = -K\hat{x}_n(t), \qquad t \geq 0 \tag{2.169}$$

is designed, in which (C_n, A_n, B_n) are the parameters of the approximation Σ_n with the dominant eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}^+_{-\beta}$. Since it is the basic concept of the

early-lumping approach to consider only the approximation for the compensator design instead of the infinite-dimensional system, the intention to achieve the stability margin β for the closed-loop system requires the eigenvalues of the controlled approximation to be located in $\overline{\mathbb{C}}^-_{-\beta}$. It is a well-known result that these coincide with the eigenvalues of the matrices $A_n - B_n K$ and $A_n - LC_n$ (see [86, Sec. 4.2]). These eigenvalues can be assigned by standard techniques for the computation of appropriate gains K and L. The eigenvalues $\lambda_{n+1}, \lambda_{n+2}, \ldots$ of the residual dynamics cannot be influenced by this control method in a systematic way. But since these are assumed to be contained in $\overline{\mathbb{C}}^-_{-\beta}$ anyway (see (2.164)) they can be left unchanged by the control. Thus, the union of the assigned spectra $\sigma(A_n - B_n K)$ and $\sigma(A_n - LC_n)$ as well as the unmodified spectrum $\sigma(\mathcal{A}_r)$ of the residual dynamics is regarded as the desired spectrum of the closed-loop system. As explained in the previous subsection Assumption 2.3-1 guarantees that the pair (A_n, B_n) is controllable and the pair (C_n, A_n) is observable. Thus, the *controller gain* $K \in \mathbb{C}^{p\times n}$ and the *observer gain* $L \in \mathbb{C}^{n\times m}$ can be computed such that $\sigma(A_n - B_n K) \cup \sigma(A_n - LC_n) \subset \overline{\mathbb{C}}^-_{-\beta}$, for which purpose standard methods for finite-dimensional LTI systems can be applied.

The impact of neglecting the residual dynamics for the compensator design becomes apparent when it is noted that not the output $y_n(t)$ of the approximation is used in (2.168) but instead the output $y(t)$ of the infinite-dimensional system. This comes from the fact that $y_n(t)$ is not available by measurement and is legitimated by the assumption that the approximation describes the dominant part of the system's transfer behavior so that the error $y(t) - y_n(t) = y_r(t)$ can be ignored. The impact of this disregard becomes apparent when the dynamics of the *observation error* $e_n(t) \coloneqq x_n(t) - \hat{x}_n(t)$, that are referred to as *observer dynamics*, are considered that are given by

$$\dot{e}_n(t) = (A_n - LC_n)e_n(t) - L\mathcal{C}_r x_r(t) \tag{2.170}$$

in view of (2.99) and (2.168), using (2.100), (2.118), and (2.120). This equation shows that the observer dynamics are excited by the residual state x_r, an effect that is introduced in [5] as observation spillover.

If the residual state x_r approaches zero, also the error e_n will vanish exponentially since $A_n - LC_n$ is a Hurwitz matrix. However, the residual dynamics are excited by e_n in view of the control law $u(t) = -K\hat{x}_n(t) = -Kx_n(t) + Ke_n(t)$, yielding

$$\dot{x}_r(t) = \mathcal{A}_r x_t(t) - \mathcal{B}_r K x_n(t) + \mathcal{B}_r K e_n(t) \tag{2.171}$$

(see (2.117) and (2.169)). This influence on the residual dynamics is called *control*

spillover (see [5]). Since the residual dynamics and the observer dynamics impact each other in both directions the control performance may be deteriorated which even can lead to instability. This interplay of observation and control spillover is analyzed in the next subsection.

2.3.3 Dynamics of the closed-loop system

Since the residual dynamics are not taken into account for the compensator design, the impact of the compensator on the actual infinite-dimensional system Σ is not clear a priori. Therefore, the dynamics of the closed-loop system will be analyzed next. To this end, the composed state

$$x_{cl}(t) := \begin{bmatrix} e_n(t) \\ x_n(t) \\ x_r(t) \end{bmatrix} \tag{2.172}$$

with $e_n(t) = x_n(t) - \hat{x}_n(t)$ on the state space $X_{cl} := \mathbb{C}^n \oplus \mathbb{C}^n \oplus X_r$ is considered that is a Hilbert space with the inner product

$$\left\langle \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \right\rangle_{X_{cl}} := \langle g_1, h_1 \rangle_{\mathbb{C}^n} + \langle g_2, h_2 \rangle_{\mathbb{C}^n} + \langle g_3, h_3 \rangle_{X_r} \tag{2.173}$$

with g_1, g_2, h_1, $h_2 \in \mathbb{C}^n$ and g_3, $h_3 \in X_r$, where $\langle g_1, h_1 \rangle_{\mathbb{C}^n} = \sum_{i=1}^{n} g_{1,i} \overline{h_{1,i}}$ is the usual inner product on $\mathbb{C}^n$. It can be verified easily by insertion of (2.169) into (2.99) and (2.117) in combination with (2.170) that the closed-loop dynamics are described by the abstract initial value problem

$$\Sigma_{cl}: \quad \dot{x}_{cl}(t) = \mathcal{A}_{cl} x_{cl}(t), \qquad t > 0, \quad x_{cl}(0) = x_{cl,0} \in X_{cl} \tag{2.174}$$

with the *closed-loop system operator*

$$\mathcal{A}_{cl} = \begin{bmatrix} A_n - LC_n & 0 & -L\mathcal{C}_r \\ B_n K & A_n - B_n K & 0 \\ \mathcal{B}_r K & -\mathcal{B}_r K & \mathcal{A}_r \end{bmatrix} \tag{2.175}$$

with $D(\mathcal{A}_{cl}) = \mathbb{C}^n \oplus \mathbb{C}^n \oplus D(\mathcal{A}_r) \subset X_{cl}$. For the analysis of the spillover this operator is decomposed according to

$$\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta, \tag{2.176}$$

wherein

$$\mathcal{A}_{cl,0} = \begin{bmatrix} A_n - LC_n & 0 & 0 \\ B_n K & A_n - B_n K & 0 \\ \mathcal{B}_r K & -\mathcal{B}_r K & \mathcal{A}_r \end{bmatrix}, \qquad \Delta = \begin{bmatrix} 0 & 0 & -L\mathcal{C}_r \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.177}$$

with $D(\mathcal{A}_{cl,0}) = D(\mathcal{A}_{cl})$ and $D(\Delta) = X_{cl}$. From the fact that $A_n - LC_n$, $A_n - B_nK$, and $\mathcal{A}_r$ generate C_0-semigroups and from the boundedness of the off-diagonal blocks of $\mathcal{A}_{cl,0}$ it can be concluded that also the composite operator $\mathcal{A}_{cl,0}$ is the infinitesimal generator of a C_0-semigroup (see [46, Lem. 3.2.2]). Furthermore, due to the boundedness of Δ, also $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$ is an infinitesimal generator of a C_0-semigroup so that the initial value problem (2.174) is well-posed (see [46, Thm. 3.2.1]).

Before the closed-loop behavior is discussed by considering the spectrum of $\mathcal{A}_{cl}$ it has to be noted that this makes sense only if the closed-loop spectrum is meaningful for the behavior of the control loop in the way that the stability margin β is given by $\beta = -\sup_{\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{cl})} \operatorname{Re} \tilde{\lambda}_{cl}$. Therefore, the following assumption has to hold (see also Subsection 2.1.3 for the SDGA).

Assumption 2.3-2 (SDGA for the closed-loop system)

The closed-loop system operator $\mathcal{A}_{cl}$ is assumed to satisfy the spectrum determined growth assumption, *i.e.*, $\beta = -\sup_{\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{cl})} \operatorname{Re} \tilde{\lambda}_{cl}$ holds for the closed-loop stability margin β. ◀

An easy method to conclude that this assumption holds is available if $\mathcal{A}$ generates an analytic C_0-semigroup, which implies that $\mathcal{A}_r = \mathcal{A}|_{D(\mathcal{A}) \cap X_r}$ (see (2.115)) is the generator of an analytic C_0-semigroup. Since X_{cl} and X_r differ by the finite-dimensional space $\mathbb{C}^{2n}$ it can be shown that also the C_0-semigroup, generated by $\mathcal{A}_{cl,0}$, is analytic. Due to the fact that Δ is bounded, finally $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$ generates an analytic C_0-semigroup (see [89, Cor. IX 2.5]) which in turn assures that $\mathcal{A}_{cl}$ satisfies the SDGA (see [126, Sec. 2]) so that Assumption 2.3-2 holds. However, for $\mathcal{A}$ to generate an analytic C_0-semigroup it is necessary that its spectrum satisfies the sector condition (2.37) with a sector angle $\varepsilon > 0$. This excludes, *e.g.*, entirely undamped flexible systems and beam systems with viscous or Rayleigh damping. The following statement assures the SDGA for the closed-loop system for spectra of $\mathcal{A}$ that are contained in a sector with sector angle $\varepsilon = 0$.

Theorem 2.3-3

Suppose that $\mathcal{A}$ is the infinitesimal generator of a C_0-semigroup and has an eigenvector Riesz basis. Then, Assumption 2.3-2 is satisfied.

The proof is given in Appendix A.3. Apparently, this statement can be applied particularly when $\mathcal{A}$ is a sectorial Riesz-spectral operator. For proving the statement it plays an essential role that $\mathcal{A}_{cl}$ has "almost" an eigenvector Riesz basis under the mentioned condition in the following sense: $\mathcal{A}_{cl}$ has an infinite number of eigenvectors $\tilde{\phi}_{cl,i}$, $i > N$, that, when combined with an at most finite number N of suitable additional vectors $\tilde{\varphi}_i \in X_{cl}$, $i = 1, 2, \ldots, N$, yield a Riesz basis for X_{cl}. If $\mathcal{A}$ neither is the generator of an analytic C_0-semigroup nor has an eigenvector Riesz basis, the SDGA for $\mathcal{A}_{cl}$ has to be checked individually. This is demonstrated next for the control of the time-delay system that has been introduced in Example 2.1-16.

Example 2.3-4 (Time-delay system)

Applying the observer-based compensator Σ_c in (2.168)–(2.169) to the time-delay system considered in the Examples 2.1-16 and 2.2-4, the closed-loop dynamics is described by

$$\begin{bmatrix} \dot{x}_l(t) \\ \dot{\hat{x}}_n(t) \\ \partial_t x_d(\tau, t) \end{bmatrix} = \begin{bmatrix} A_l & -B_l K & A_d(\cdot)|_{\tau=-T} \\ LC_l & A_n - B_n K - LC_n & 0 \\ 0 & 0 & \partial_\tau(\cdot) \end{bmatrix} \begin{bmatrix} x_l(t) \\ \hat{x}_n(t) \\ x_d(\tau, t) \end{bmatrix} \tag{2.178}$$

with $x_d(0, t) = x_l(t)$, $t > 0$. This system has the same form as the time-delay system (2.78) with (2.76), at which the lumped part x_l of the state in (2.78) is extended by the state $\hat{x}_n$ of the compensator. For these systems it is known that the SDGA is satisfied in general (see [46, Thm. 5.1.7]). Thus, Assumption 2.3-2 holds. ◀

The spectrum of $\mathcal{A}_{cl,0}$ is given by

$$\sigma(\mathcal{A}_{cl,0}) = \sigma(A_n - B_n K) \cup \sigma(A_n - LC_n) \cup \sigma(\mathcal{A}_r) \tag{2.179}$$

due to its triangular structure. This makes apparent that the compensator, which is designed by eigenvalue assignment for $A_n - B_n K$ and $A_n - LC_n$ assigns the desired eigenvalues to $\mathcal{A}_{cl,0}$. However, the actual closed-loop dynamics are not described by $\mathcal{A}_{cl,0}$ but instead by the perturbed operator $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$ (see (2.174)). In general, all closed-loop eigenvalues $\tilde{\lambda}_{cl,i} \in \sigma(\mathcal{A}_{cl})$, $i \in \mathbb{N}$, differ therefore from the desired values $\lambda_{cl,i} \in \sigma(\mathcal{A}_{cl,0})$, $i \in \mathbb{N}$, so that the performance of the controlled system may be unsatisfying. If one of the eigenvalues is shifted in this way into the right half of the complex plane, the closed-loop system is unstable. It is therefore important to have an estimate how far the eigenvalues $\tilde{\lambda}_{cl,i}$ may be shifted away from $\lambda_{cl,i}$ under the influence of the perturbation Δ. This question will be addressed in Section 2.4.

2.3.4 Asymptotic disturbance rejection

A stabilizing control has the effect that the plant output y approaches zero in the considered structure of the control loop as it is shown in Figure 12a. For a two-degrees-of-freedom control (see Figure 12b) one has $y(t) \to y_d(t)$ for $t \to \infty$, which means that the control assures that the output approaches the desired trajectory. However, this intended behavior of the closed-loop system changes when an exogenous disturbance $d(t) \in \mathbb{R}^{p_d}$ affects the closed-loop system, *i.e.*, the plant is described by

$$\Sigma: \quad \dot{x}(t) = \mathcal{A}x(t) + \mathcal{B}u(t) + \mathcal{G}d(t), \qquad t > 0, \quad x(0) = x_0 \in X \tag{2.180}$$

$$y(t) = \mathcal{C}x(t), \qquad t \geq 0 \tag{2.181}$$

instead of the model (2.3)–(2.4) with a bounded linear operator $\mathcal{G} : \mathbb{C}^{p_d} \mapsto X$ and m outputs, *i.e.*, $y(t) \in \mathbb{R}^m$. For modifying the compensator such that it rejects the influence of the disturbance d on the output y asymptotically it is assumed that d can be described by the *signal model*

$$\dot{x}_s(t) = A_s x_s(t), \qquad t > 0, \quad x_s(0) = x_{s,0} \in \mathbb{C}^{n_s} \tag{2.182}$$

$$d(t) = C_s x_s(t), \qquad t \geq 0 \tag{2.183}$$

with (C_s, A_s) being observable, which is possible, *e.g.*, for constant, ramp, or sinusoidal signals. In the important case of constant disturbances one has $A_s = 0$, $C_s = 1$, and $n_s = 1$. An approach to robustly and asymptotically reject the modeled disturbance is the *internal model principle* (see [48, 65]). For single-output systems, *i.e.*, $m = 1$, the plant model (2.180)–(2.181) is extended by a copy

$$\dot{\tilde{x}}_s(t) = A_s \tilde{x}_s(t) + B_s y(t), \qquad \tilde{x}_s(0) = 0 \tag{2.184}$$

of the signal model that is excited by the plant output y, where $B_s \in \mathbb{C}_s^n$ is arbitrary except that (A_s, B_s) has to be controllable. For multi-output systems, *i.e.*, $m > 1$, one has to add such a copy of the signal model for each of the outputs, and $p \geq m$ has to hold (see [48]). Typically, A_s has eigenvalues on the imaginary axis in order to generate persistent disturbances. If the compensator stabilizes the extended plant, it follows

$$\lim_{t \to \infty} y(t) = 0, \qquad \forall\, x_0 \in X, \forall\, x_{s,0} \in \mathbb{C}^{n_s} \tag{2.185}$$

because $\tilde{x}_s$ is excited by y (see (2.184)) so that this state would otherwise diverge due to the instability of the signal model and the controllability of (A_s, B_s). Thus, asymptotic disturbance rejection is assured. Since the rejection is achieved whenever the control

loop is exponentially stable, the rejection is robust w.r.t. model uncertainties that do not destabilize the closed-loop system.

The design of the compensator is based on an approximation of the extended system (2.180)–(2.181) and (2.184). This approximation can be obtained by adding the signal model dynamics (2.184) to the previously used modal approximation Σ_n (see (2.99)–(2.100) and (2.138)–(2.141)), for what the output $y(t) = \mathcal{C}x(t)$ in (2.184) is approximated by $y_n(t) = C_n x_n(t)$. This leads to the approximation

$$\Sigma_{n,e}: \quad \begin{bmatrix} \dot{x}_n(t) \\ \dot{\tilde{x}}_s(t) \end{bmatrix} = \begin{bmatrix} A_n & 0 \\ B_s C_n & A_s \end{bmatrix} \begin{bmatrix} x_n(t) \\ \tilde{x}_s(t) \end{bmatrix} + \begin{bmatrix} B_n \\ 0 \end{bmatrix} u(t) \tag{2.186}$$

$$\begin{bmatrix} x_n(0) \\ \tilde{x}_s(0) \end{bmatrix} = \begin{bmatrix} x_0 \\ 0 \end{bmatrix} \in X \oplus \mathbb{C}^{n_s} \tag{2.187}$$

$$y_n(t) = \begin{bmatrix} C_n & 0 \end{bmatrix} \begin{bmatrix} x_n(t) \\ \tilde{x}_s(t) \end{bmatrix}. \tag{2.188}$$

In this approximation the disturbance d is not taken into account because it does neither influence the compensator design nor the closed-loop spectrum. The state feedback (2.169) can now be designed for the extended approximation, where this (unstable) dynamics can be stabilized if and only if the approximation Σ_n does not have any transmission zero that coincides with any of the eigenvalues of A_s in view of the assumed controllability of (A_n, B_n) and (A_s, B_s) (see [48]). Since the signal model (2.184) has to be implemented, the state $\tilde{x}_s$ is directly available for realizing the state feedback and needs not to be reconstructed by the observer. Thus, the observer (2.168) can still be designed for the unextended approximation Σ_n. This leads to the compensator

$$\Sigma_{c,e}: \ \dot{\hat{x}}_n(t) = (A_n - LC_n)\hat{x}_n(t) + B_n u(t) + Ly(t), \quad t > 0, \ \hat{x}_n(0) = \hat{x}_{n,0} \in \mathbb{C}^n \tag{2.189}$$

$$u(t) = -K\tilde{x}_n(t) - K_s\tilde{x}_s(t), \qquad t \geq 0, \tag{2.190}$$

where $\tilde{x}_s$ is taken from (2.184), which has to be implemented additionally. Apparently, this control scheme differs from that considered in the previous subsection for the analysis of the closed-loop dynamics. However, it will be shown later that the spillover reduction approaches, presented in the Chapters 3 and 5, can be applied to both the control with and without asymptotic disturbance rejection (see the Remarks 2.4-14 and 4.2-5). It should be noted that the internal model principle may lead to severe *windup problems* in case of input saturation. Therefore, windup prevention measures are necessary that can be found in [51].

2.4 Analysis of the closed-loop spectrum

Different from the finite-dimensional case the dynamical behavior of the closed-loop system depends not only on the eigenvalues of the system operator but also on its continuous spectrum $\sigma_c(\mathcal{A}_{cl})$ as well as the residual spectrum $\sigma_r(\mathcal{A}_{cl})$ (for their definitions see Subsection 2.1.3). For that reason the structure of $\sigma(\mathcal{A}_{cl})$, *i.e.*, the decomposition $\sigma(\mathcal{A}_{cl}) = \sigma_p(\mathcal{A}_{cl}) \cup \sigma_c(\mathcal{A}_{cl}) \cup \sigma_r(\mathcal{A}_{cl})$, has to be analyzed. This will be discussed in the following subsection. Afterwards an estimate of the eigenvalue perturbation caused by the spillover is given in Subsection 2.4.2.

2.4.1 Structure of the closed-loop spectrum

While the spectrum of $\mathcal{A}_{cl,0}$ can be described by the simple expression (2.179), the spectrum structure of the actual closed-loop system operator $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$ is far less obvious due to the influence of the perturbation operator Δ. It is useful to note, however, that Δ is a *degenerate operator*, *i.e.*, Δ is bounded and its range $\operatorname{ran}\Delta$ is finite-dimensional (see (2.177)), because this simplifies to gain some insight into the structure of the closed-loop spectrum $\sigma(\mathcal{A}_{cl})$. One consequence is that, if $\sigma(\mathcal{A}_r)$ does not contain any accumulation point, the spectrum of $\mathcal{A}_{cl}$ entirely consists of isolated eigenvalues with finite multiplicities and thus is discrete. This can be shown by aid of the *first Weinstein-Aronszajn formula* (see [89, Sec. IV 6]). The structure analysis of the closed-loop spectrum is more sophisticated if $\sigma(\mathcal{A})$ contains accumulation points. An example for such systems are mechanical vibrating systems with *Kelvin-Voigt damping* (see, *e.g.*, [73, Sec. 4]). It is a known fact that an accumulation point $\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{cl})$ cannot be shifted by any degenerate perturbation operator and thus not by the control law (2.169) so that the stability margin β is limited by the location of the accumulation points in $\sigma(\mathcal{A})$ (see, *e.g.*, [46, Sec. 5.2]). However, the following questions need some further analysis:

1. Under which circumstances may $\sigma_c(\mathcal{A}_{cl})$ or $\sigma_r(\mathcal{A}_{cl})$ be non-empty?
2. May it happen that $\sigma_c(\mathcal{A}_{cl})$ or $\sigma_r(\mathcal{A}_{cl})$ contain elements that do not belong to $\overline{\sigma_p(\mathcal{A}_{cl})}$?
3. Can it happen that accumulation points in $\sigma(\mathcal{A})$ appear in $\sigma(\mathcal{A}_{cl})$ as isolated points instead of accumulation points?

Particularly, Item 2 is an important issue because if the closed-loop can possess a spectral point $\tilde{\lambda}_{cl} \in \sigma_c(\mathcal{A}_{cl}) \cup \sigma_r(\mathcal{A}_{cl})$ that is not in the closure of $\sigma_p(\mathcal{A}_{cl})$, it would not be sufficient to analyze only the closed-loop eigenvalues. A detailed characterization of the closed-loop spectrum $\sigma(\mathcal{A}_{cl})$ can be derived from the following general result for the spectrum of a perturbed operator.

Lemma 2.4-1

Suppose that a linear operator $\mathcal{M}_0 : D(\mathcal{M}_0) \subset H \mapsto H$ on a Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ has the following properties:

(P1) $\mathcal{M}_0$ is the generator of a C_0-semigroup,

(P2) the point spectrum $\sigma_p(\mathcal{M}_0)$ of $\mathcal{M}_0$ consists solely of isolated eigenvalues with finite algebraic multiplicities,

(P3) $\sigma(\mathcal{M}_0)$ is totally disconnected, and

(P4) the residual spectrum satisfies $\sigma_r(\mathcal{M}_0) = \emptyset$.

Let the perturbation operator $\mathcal{D} : H \mapsto H$ be degenerate. Then, the spectrum of $\mathcal{M} = \mathcal{M}_0 + \mathcal{D}$ satisfies the following statements.

(S1) $\sigma_c(\mathcal{M}) \subseteq \sigma_c(\mathcal{M}_0) \subset \sigma(\mathcal{M})$

(S2) $\sigma_r(\mathcal{M}) = \emptyset$

(S3) Any spectral point λ of $\mathcal{M}$ in the set $\sigma(\mathcal{M}) \setminus \sigma_c(\mathcal{M}_0)$ is an isolated eigenvalue of $\mathcal{M}$ with finite algebraic multiplicity.

Furthermore, it holds

$$\sigma(\mathcal{M}) = \overline{\sigma_p(\mathcal{M})} \tag{2.191}$$

so that $\sigma_c(\mathcal{M})$ entirely consists of accumulation points of eigenvalues in $\sigma_p(\mathcal{M})$.

For the proof see Appendix A.4. The first Item (S1) states that the accumulation points in $\mathcal{M}_0$ are not influenced by the perturbation und are thus also accumulation points in $\mathcal{M}$, and the Properties (S2) and (S3) show that all the remaining spectral points of $\mathcal{M}$ are eigenvalues with finite algebraic multiplicities. Note, that $\sigma_c(\mathcal{M}) = \sigma_c(\mathcal{M}_0)$ is not assured in general because an element of $\sigma_c(\mathcal{M}_0)$ may appear in $\sigma_p(\mathcal{M})$ as an accumulation point that is an eigenvalue. Instead, only $\sigma_c(\mathcal{M}) \subseteq \sigma_c(\mathcal{M}_0)$ is assured.

Remark 2.4-2 In the case where $\mathcal{M}_0$ satisfies the additional requirement to be the infinitesimal generator of an analytic C_0-semigroup the statements of Lemma 2.4-1 hold even for unbounded perturbations $\mathcal{D}$ that are $\mathcal{M}_0$-*degenerate*. Such an operator $\mathcal{D} : D(\mathcal{D}) \subset H \mapsto H$ with $D(\mathcal{D}) \supseteq D(\mathcal{M}_0)$ has a finite-dimensional range $\operatorname{ran}\mathcal{D}$ and is *unbounded* in H but $\mathcal{M}_0$-*bounded*, *i.e.*, $\|\mathcal{D}h\|_H \leq \alpha\|\mathcal{M}_0 h\|_H + \gamma\|h\|_H$, $\forall\, h \in D(\mathcal{D})$, holds for constants $\alpha, \gamma \geq 0$. In order to show this it plays an essential role that the infinitesimal generator of an analytic C_0-semigroup remains such a generator under $\mathcal{M}_0$-degenerate perturbations, which follws from [133, Prop. 1] and [89, Rem. IV 1.13]. This generalization is important for modelling pointwise measurements of parabolic systems (see, *e.g.*, [52]). ◀

Now, this general result shall be applied in order to characterize the structure of the closed-loop spectrum. To this end, remember that it has been argued in Subsection 2.3.3 that $\mathcal{A}_{cl,0}$ has the Property (P1) in Lemma 2.4-1. This is implied by the fact that $\mathcal{A}$ is the infinitesimal generator of a C_0-semigroup due to Item 1 of Assumption 2.1-9. Furthermore, the Properties (P2)–(P4) can be verified for $\mathcal{A}_{cl,0}$ under use of the Items 3a, 3b, and 3c of Assumption 2.1-9 and (2.179). Thus, $\mathcal{A}_{cl,0}$ satisfies the requirements on $\mathcal{M}_0$ in Lemma 2.4-1, and Δ is a degenerate operator as follows from (2.177) so that the above result can be applied to $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$. Note, that for doing so it is not required that $\mathcal{A}_{cl,0}$ has simple eigenvalues or an eigenvector Riesz basis. Apparently, (2.179) shows that $\sigma_c(\mathcal{A}_{cl,0}) = \sigma_c(\mathcal{A}_r) = \sigma_c(\mathcal{A})$. Hence, Lemma 2.4-1 directly yields the following characterization of the closed-loop spectrum, which is the main result of this section.

Theorem 2.4-3

Let the Assumption 2.1-9 hold. Then, the spectrum $\sigma(\mathcal{A}_{cl})$ of the closed-loop system operator $\mathcal{A}_{cl}$ can be decomposed as

$$\sigma(\mathcal{A}_{cl}) = \{\tilde{\lambda}_{cl,i}, i \in \mathbb{N}\} \cup \sigma_c(\mathcal{A}), \tag{2.192}$$

where $\tilde{\lambda}_{cl,i}$, $i \in \mathbb{N}$, are eigenvalues of $\mathcal{A}_{cl}$ that have finite algebraic multiplicities and are isolated. Particularly,

$$\sigma_r(\mathcal{A}_{cl}) = \emptyset \tag{2.193}$$

$$\sigma(\mathcal{A}_{cl}) = \overline{\sigma_p(\mathcal{A}_{cl})} \tag{2.194}$$

holds.

Apparently, any spectral point in $\sigma(\mathcal{A}_{cl})$ that is not an eigenvalue is an accumulation point in $\sigma(\mathcal{A})$ and in $\sigma(\mathcal{A}_{cl})$. However, an accumulation point in $\sigma(\mathcal{A}_{cl})$ might be an eigenvalue with infinite multiplicity. This result greatly simplifies the analysis of the closed-loop dynamics since the accumulation points in $\sigma(\mathcal{A})$ are usually known so that one needs only to determine the closed-loop eigenvalues. An approach for estimating them is shown in the next subsection. It is interesting to note that the same result as in Theorem 2.4-3 would be obtained from Lemma 2.4-1 in the more general case where instead of the observer-based state feedback (2.168)–(2.169) an arbitrary finite-dimensional compensator for the system (2.3)–(2.4) is used. This follows from the fact that for *any* controller the closed-loop system operator $\mathcal{A}_{cl}$ differs from the open-loop system operator $\mathcal{A}_{cl,0}$ by a degenerate perturbation.

According to Theorem 2.4-3 the closed-loop spectrum equals the closure of its point spectrum (see (2.194)). In order to determine the closed-loop stability margin under Assumption 2.3-2 it is therefore sufficient to check only the eigenvalues instead of the whole spectrum. Thus, the following statement is obvious.

Corollary 2.4-4

Let the Assumptions 2.1-9 and 2.3-2 hold. Then,

$$\omega_{cl} = \sup_{\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{cl})} \operatorname{Re} \tilde{\lambda}_{cl} = \sup_{\tilde{\lambda}_{cl} \in \sigma_p(\mathcal{A}_{cl})} \operatorname{Re} \tilde{\lambda}_{cl} \tag{2.195}$$

is the growth bound ω_{cl} of the C_0-semigroup $\mathcal{S}_{cl}(t)$ generated by $\mathcal{A}_{cl}$.

This result confirms that by shifting the dominant eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathcal{A}$ such that all the eigenvalues $\tilde{\lambda}_{cl,i}$, $i \in \mathbb{N}$, of $\mathcal{A}_{cl}$ are contained in the half-plane $\overline{\mathbb{C}}^{-}_{-\beta}$, the stability margin β is achieved, as it is well-known from the finite-dimensional case.

Example 2.4-5 (Euler-Bernoulli beam with Kelvin-Voigt damping, continued)

The Euler-Bernoulli beam with Kelvin-Voigt damping introduced in Example 2.1-15 is considered. There, it has been found that $\mathcal{A}$ is a sectorial Riesz-spectral operator so that the SDGA for the closed-loop system holds because $\mathcal{A}$ is analytic (see the explanation following Assumption 2.3-2). Furthermore, it has been stated that the continuous spectrum consists of the accumulation point $\lambda^{acc} = -1/(2\delta)$, *i.e.*, $\sigma_c(\mathcal{A}) = \{\lambda^{acc}\}$. In addition, it has been explained that the corresponding state space model

satisfies Assumption 2.1-9 so that Theorem 2.4-3 can be applied to this system. Thus, according to (2.192) the closed-loop spectrum $\sigma(\mathcal{A}_{cl})$ consists of isolated eigenvalues with finite multiplicities and the accumulation point λ^{acc}. This means in particular that this spectral point cannot be shifted by the compensator and that $\sigma_c(\mathcal{A}_{cl})$ can contain at most λ^{acc}. Further, $\sigma_r(\mathcal{A}_{cl}) = \emptyset$ holds. Since $\mathcal{A}_{cl}$ satisfies the SDGA, *i.e.*, Assumption 2.3-2 holds, Corollary 2.4-4 can be applied which states that it is sufficient to consider the point spectrum of $\mathcal{A}_{cl}$ in order to determine the growth bound of the corresponding C_0-semigroup. ◀

Clearly, a more specific characterization of the closed-loop behavior can be obtained when the locations of the eigenvalues $\tilde{\lambda}_{cl,i}$ are known. However, since the eigenvalues considered for the compensator design are the eigenvalues $\lambda_{cl,i}$, $i \in \mathbb{N}$, of $\mathcal{A}_{cl,0}$ due to (2.179), the difference between $\lambda_{cl,i}$ and $\tilde{\lambda}_{cl,i}$ has to be taken into account in the design step. An upper bound for these errors is obtained in the next subsection.

2.4.2 Enclosure of the closed-loop spectrum

Now, a relation between the perturbed spectrum $\sigma(\mathcal{A}_{cl})$ of $\mathcal{A}_{cl}$ and the spectrum $\sigma(\mathcal{A}_{cl,0})$ of the nominal operator $\mathcal{A}_{cl,0}$ will be determined. To be more precise, the *spectrum perturbation*

$$d := \sup_{\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{cl})} \inf_{\lambda_{cl} \in \sigma(\mathcal{A}_{cl,0})} |\tilde{\lambda}_{cl} - \lambda_{cl}| \tag{2.196}$$

will be estimated by an upper bound. Note, that the term $\inf_{\lambda_{cl} \in \sigma(\mathcal{A}_{cl,0})} |\tilde{\lambda}_{cl} - \lambda_{cl}|$ therein describes the distance of the spectral point $\tilde{\lambda}_{cl}$ of the perturbed spectrum to the set $\sigma(\mathcal{A}_{cl,0})$. Thus, d can be regarded as the largest distance of a point $\tilde{\lambda}_{cl}$ to its neighboring spectral point in the unperturbed spectrum $\sigma(\mathcal{A}_{cl,0})$. Knowing this worst-case difference enables to estimate not only the stability margin of the closed-loop system but also further performance aspects such as its damping. Therefore, having an estimate for d, the early-lumping-based compensator design approach by assigning the eigenvalues of the controlled approximation is justified since the closed-loop behavior can be estimated *a priori*.

A result for the spectrum perturbation can be obtained by a generalized version of the *Bauer-Fike Theorem* [18, Thm. IIIa] which provides estimates for the eigenvalues of a finite-dimensional matrix under bounded perturbations. For the purposes of this

contribution an extension to a class of unbounded operators is used that are similar to a *normal operator*[33]. This generalized statement reads as follows.

Lemma 2.4-6

Suppose that for a linear operator $\mathcal{M}_0 : D(\mathcal{M}_0) \subset H \mapsto H$ *on a Hilbert space* $(H, \langle\cdot,\cdot\rangle_H)$ *there exists a linear transformation*[34] $\mathcal{T} : H \mapsto H$ *such that* $\mathcal{T}^{-1}\mathcal{M}_0\mathcal{T}$ *is normal. In addition, let* $\mathcal{D} : H \mapsto H$ *be a bounded operator. If the operator* $\mathcal{M} = \mathcal{M}_0 + \mathcal{D}$ *satisfies* $\sigma_r(\mathcal{M}) = \emptyset$*, then the spectrum perturbation*

$$d_{\mathcal{M}} := \sup_{\tilde{\lambda}\in\sigma(\mathcal{M})} \inf_{\lambda\in\sigma(\mathcal{M}_0)} |\tilde{\lambda} - \lambda| \tag{2.199}$$

with respect to $\mathcal{M}_0$ *and* $\mathcal{M}$ *satisfies*

$$d_{\mathcal{M}} \leq \|\mathcal{T}^{-1}\mathcal{D}\mathcal{T}\|. \tag{2.200}$$

The proof is given in Appendix A.5. Note, that different from Lemma 2.4-1 this statement cannot be generalized in a straightforward manner to perturbation operators D that are $\mathcal{M}_0$-bounded. However, this result is applied next for analyzing the spectrum of $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$, where the perturbation operator Δ satisfies the condition to be bounded (see (2.177)). For doing so, the following assumption is needed.

Assumption 2.4-7 (Existence of a normalizing transformation)

The existence of a linear transformation $\mathcal{T}_{cl} : X_{cl} \mapsto X_{cl}$, such that $\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl}$ is normal, is assumed. ◀

This condition can be assured in a particular simple way if the system operator is a Riesz-spectral operator. For this case the following statement is useful.

[33] A linear operator $\mathcal{M} : X \mapsto X$ is said to be *normal* if it holds

$$\mathcal{M}\mathcal{M}^*h = \mathcal{M}^*\mathcal{M}h, \qquad \forall\, h \in D(\mathcal{M}\mathcal{M}^*) \tag{2.197}$$

$$D(\mathcal{M}\mathcal{M}^*) = D(\mathcal{M}^*\mathcal{M}). \tag{2.198}$$

[34] A linear operator $\mathcal{T} : X \mapsto X$ is called a *linear transformation* if it is a bijection on whole X. Thus, $D(\mathcal{T}) = D(\mathcal{T}^{-1}) = X$ which implies that $\mathcal{T}$ and $\mathcal{T}^{-1}$ are bounded.

Proposition 2.4-8

Suppose that $\mathcal{A}$ is a Riesz-spectral operator. If the observer-based compensator Σ_c in (2.168)–(2.169) is designed such that $A_n - B_nK$ and $A_n - LC_n$ have simple and mutually different eigenvalues that are not contained in $\sigma(\mathcal{A}_r)$, then Assumption 2.4-7 holds. In particular, a normalizing transformation is then given by

$$\mathcal{T}_{cl}h := \sum_{i=1}^{\infty} \langle h, \varphi_i \rangle_X \phi_{cl,i}, \qquad \mathcal{T}_{cl}^{-1}h = \sum_{i=1}^{\infty} \langle h, \psi_{cl,i} \rangle_X \varphi_i, \qquad \forall\, h \in X, \tag{2.201}$$

wherein $\phi_{cl,i}$ and $\psi_{cl,i}$, $i \in \mathbb{N}$, are the eigenvectors of $\mathcal{A}_{cl,0}$ and $\mathcal{A}_{cl,0}^*$, respectively, that correspond to each other and are scaled such that they form biorthonormal sequences. Furthermore, φ_i, $i \in \mathbb{N}$, is an arbitrary orthonormal basis for X_{cl}.

For the proof see Appendix A.6.

Remark 2.4-9 This result remains valid if $\mathcal{A}$ is not a Riesz-spectral operator but the eigenvectors $\phi_{cl,i}$ of the closed-loop system operator $\mathcal{A}_{cl,0}$ form a Riesz basis, and $(\psi_{cl,i})_{i\in\mathbb{N}}$ is the biorthonormal sequence corresponding to $(\phi_{cl,i})_{i\in\mathbb{N}}$. This is relevant for systems with multiple eigenvalues whose system operator is therefore not Riesz-spectral in the sense of Definition 2.1-6 (see [73] for a more general definition). Furthermore, Assumption 2.4-7 is satisfied for system operators $\mathcal{A}$ that are self-adjoint, skew-adjoint, or unitary, even if their eigenvectors do not form a Riesz basis. In this case, a normalizing transformation is obtained from (2.201), when $\phi_{cl,1}, \phi_{cl,1}, \ldots, \phi_{cl,2n}$ are the eigenvectors of $\mathcal{A}_{cl,0}$ that correspond to the eigenvalues of $A_n - B_nK$ and $A_n - LC_n$, the vectors $\phi_{cl,i}$, $i > 2n$, and φ_i, $i \in \mathbb{N}$, are chosen as

$$\varphi_i = \begin{cases} \begin{bmatrix} e_i & 0 & 0 \end{bmatrix}^T & : \quad 1 \le i \le n \\ \begin{bmatrix} 0 & e_{i-n} & 0 \end{bmatrix}^T & : \quad n < i \le 2n \\ \begin{bmatrix} 0 & 0 & \varphi_{r,i-2n} \end{bmatrix}^T & : \quad i > 2n \end{cases}, \qquad \phi_{cl,i} = \varphi_i, \quad i > 2n \tag{2.202}$$

with e_i, $i = 1, 2, \ldots, n$, and $\varphi_{r,i}$, $i \in \mathbb{N}$, being the canonical basis for $\mathbb{R}^n$ and an orthonormal basis for X_r, respectively, and $(\psi_{cl,i})_{i\in\mathbb{N}}$ is the biorthonormal sequence corresponding to $(\phi_{cl,i})_{i\in\mathbb{N}}$. However, for time-delay systems as considered in Example 2.1-16 a normalizing transformation is easy to determine only under restrictive conditions (see [46, Sec. 2.4]), so that the following spectrum perturbation estimate cannot be applied to this system class. ◀

Now, Lemma 2.4-6 can be applied to $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$ under the Assumption 2.4-7 because all requirements are met in view of Δ being a bounded perturbation, and $\sigma_r(\mathcal{A}_{cl}) = \emptyset$ holds according to Theorem 2.4-3. Applying Lemma 2.4-6 immediately leads to the following result.

Theorem 2.4-10

Let the Assumptions 2.1-9 and 2.4-7 hold. Then,

$$d \leq \|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\| \leq \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\| \tag{2.203}$$

is an upper bound for the spectrum perturbation d, in which $\mathcal{T}_{cl}$ is the normalizing transformation in the sense of Assumption 2.4-7.

Remark 2.4-11 An equivalent formulation of this statement is that the spectrum of $\mathcal{A}_{cl}$ is enclosed by the union

$$\sigma(\mathcal{A}_{cl}) \subseteq \bigcup_{\lambda_{cl} \in \overline{\sigma_p(\mathcal{A}_{cl,0})}} \left\{ \tilde{\lambda}_{cl} \in \mathbb{C} \;\middle|\; |\tilde{\lambda}_{cl} - \lambda_{cl}| \leq r \right\} \tag{2.204}$$

of disks with radius

$$r = \|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\|, \tag{2.205}$$

at which $\sigma(\mathcal{A}_{cl,0}) = \overline{\sigma_p(\mathcal{A}_{cl,0})}$ (see Item 3d of Assumption 2.1-9) has been taken into account. Thus, Theorem 2.4-10 shows that all the eigenvalues $\tilde{\lambda}_{cl,i}$, $i \in \mathbb{N}$, are enclosed by at least one of the disks that are centered around the unperturbed eigenvalues $\lambda_{cl,i} \in \sigma(\mathcal{A}_{cl,0})$ that are known (see (2.179)). ◀

Remark 2.4-12 A similar result can be obtained by the so-called *Gerschgorin disks*. In [68] this approach was extended to the situation where the system operator can be represented as an infinite-dimensional matrix. However, in order to assure that the radii of the enclosing circles are finite the system operator has to satisfy certain conditions that are restrictive for applications with discontinuous input and output distribution functions b_i and c_i (see (2.14)–(2.15)). In contrast, the radius r in (2.205) is always finite because Δ, $\mathcal{T}_{cl}$, and $\mathcal{T}_{cl}^{-1}$ in (2.203) are bounded. Of course, it is

intuitively clear that if the controller is designed such that $\|\Delta\|$ is small, then the eigenvalue deviation caused by spillover is small as well. However, Theorem 2.4-10 provides a *quantitative* and simple relation for the possible differences between the unperturbed eigenvalues and the actual closed-loop eigenvalues. ◀

In Appendix B an approach is given for computing the radius r of the enclosure (2.204)–(2.205) exactly under the condition that $\mathcal{A}$ is a Riesz-spectral operator with orthonormal eigenvectors. In other situations r has to be computed on the basis of an accurate approximation of the closed-loop dynamics. To this end, one increases the approximation order until the resulting values for r converge.

Example 2.4-13 (Heat conducting plate, continued)

In order to demonstrate the presented approach the control of the heat conducting plate, for which a state space model has been determined in Example 2.1-13, is considered. The temperature $x(z_1, z_2, t)$ over the rectangular domain $\Omega = (0, 1) \times (0, \pi)$ with the two spatial coordinates $z_1 \in [0, 1]$ and $z_2 \in [0, \pi]$ is described by the a state space model (2.3)–(2.4) with the operators

$$\mathcal{A}h = \frac{\partial^2 h}{\partial z_1^2} + \frac{\partial^2 h}{\partial z_2^2}, \qquad \forall\, h \in D(\mathcal{A}) = \left\{ h \in H_2(\Omega) \;\middle|\; h\,\big|_\Gamma = 0 \right\} \tag{2.206}$$

$$\mathcal{B}v = b\,v, \qquad \forall\, v \in \mathbb{C} \tag{2.207}$$

$$\mathcal{C}h = \langle h, c \rangle_X, \qquad \forall\, h \in X, \tag{2.208}$$

where $\Gamma = \partial\Omega$ denotes the boundary of the plate and the input and output distribution functions are given by

$$b(z_1, z_2) = \begin{cases} 1 - (z_1 - 1/2)^2 - (z_2 - \pi/4)^2) & : \quad 0 \le z_1 \le 1,\ 0 \le z_2 \le \pi/2 \\ 0 & : \quad \text{otherwise} \end{cases} \tag{2.209}$$

$$c(z_1, z_2) = \begin{cases} 10^3 & : \quad 0.49 \le z_1 \le 0.51,\ \pi/4 - 0.01 \le z_2 \le \pi/4 + 0.01 \\ 0 & : \quad \text{otherwise.} \end{cases} \tag{2.210}$$

For convenience, the eigenvalues and eigenvectors, given in (2.45)–(2.46), are re-indexed as λ_i and ϕ_i, $i = 1, 2, \ldots$, such that $\lambda_{i+1} < \lambda_i$, $i \in \mathbb{N}$. It is apparent from (2.203) that the estimate for the spectrum perturbation d depends on the norm of the perturbation operator Δ. In the following, it will be shown that this norm can be changed by considering the transformed state $\tilde{x}(t) = \mathcal{T}_x x(t)$, which is advantageous for obtaining

tight eigenvalue enclosures. The operator $\mathcal{T}_x : X \mapsto X$ is the self-adjoint, positive, and bounded linear operator

$$\mathcal{T}_x h = \sum_{i=1}^{\infty} \rho_i \langle h, \phi_i \rangle_X \phi_i, \quad m \leq \rho_i \leq M,\ i \geq 1,\ \ \forall\, h \in X \tag{2.211}$$

with constants $m, M > 0$. $\mathcal{T}_x$ has a bounded inverse, because all its eigenvalues ρ_i satisfy $\rho_i > m > 0$, that is given by $\mathcal{T}_x^{-1} = \sum_{i=1}^{\infty} \rho_i^{-1} \langle \cdot, \phi_i \rangle_X \phi_i$. Rewriting the state space model (2.3)–(2.4) in terms of the new state $\tilde{x}(t) = \mathcal{T}_x x(t)$ leads to the transformed state space model

$$\dot{\tilde{x}}(t) = \mathcal{A}\tilde{x}(t) + \mathcal{T}_x \mathcal{B} u(t), \qquad t > 0,\ \tilde{x}(0) = \mathcal{T}_x x_0 \tag{2.212}$$

$$y(t) = \mathcal{C}\mathcal{T}_x^{-1} \tilde{x}(t), \qquad t \geq 0. \tag{2.213}$$

Observe that only the input and output operators are modified by the transformation but the system operator remains unchanged because $\mathcal{T}_x \mathcal{A} \mathcal{T}_x^{-1} = \mathcal{A}$ holds as can be checked easily. Combining (2.211) and the representations

$$\mathcal{B}v = bv = \sum_{i=1}^{\infty} \langle bv, \phi_i \rangle_X \phi_i, \qquad \forall\, v \in \mathbb{C} \tag{2.214}$$

$$\mathcal{C}h = \langle h, c \rangle_X = \left\langle h, \sum_{i=1}^{\infty} \langle c, \phi_i \rangle_X \phi_i \right\rangle_X = \sum_{i=1}^{\infty} \langle c, \phi_i \rangle_X \langle h, \phi_i \rangle_X, \qquad \forall\, h \in X \tag{2.215}$$

and $h = \sum_{i=1}^{\infty} \langle h, \phi_i \rangle_X \phi_i$ (see (2.20)), that are possible because the eigenvectors ϕ_i are orthonormal (see (2.46)), leads to

$$\tilde{\mathcal{B}}v \coloneqq \mathcal{T}_x \mathcal{B} v = \sum_{i=1}^{\infty} \rho_i \langle bv, \phi_i \rangle_X \phi_i, \qquad \forall\, v \in \mathbb{C} \tag{2.216}$$

$$\tilde{\mathcal{C}}h \coloneqq \mathcal{C}\mathcal{T}_x^{-1} h = \sum_{i=1}^{\infty} \frac{1}{\rho_i} \langle c, \phi_i \rangle_X \langle h, \phi_i \rangle_X, \qquad \forall\, h \in X. \tag{2.217}$$

Although the observer gain L is not fixed yet it is apparent from (2.217) that $\|L\mathcal{C}_r\|$ with $\mathcal{C}_r = \tilde{\mathcal{C}}|_{X_r}$ (see (2.115)) can be made arbitrarily small by choosing ρ_i, $i \geq 1$, sufficiently large. At first glance, increasing ρ_i results in reducing the radius r of the enclosures in view of (2.205) and $\|\Delta\| = \|L\mathcal{C}_r\|$ (see (2.177)). However, at the same time when $\|L\mathcal{C}_r\|$ is reduced by using large parameters ρ_i the norm $\|\mathcal{B}_r K\|$ is increased for a given feedback gain K as can be seen from (2.216). This in turn often leads to an increased norm $\|\mathcal{T}_{cl}^{-1}\|$. So, making $\|\Delta\|$ small by the state transformation yields $\|\mathcal{T}_{cl}^{-1}\|$ being large, and the other way round. In view of (2.203) one has therefore to

find a balance between $\|L\mathcal{C}_r\|$ and $\|\mathcal{B}_r K\|$ in order to obtain good estimates. To this end, the parameters ρ_i are chosen in this example as

$$\rho_1 = 1.5, \quad \rho_2 = 0.1, \quad \rho_3 = 2, \quad \rho_\infty := \rho_i = 15 \ \text{ for } \ i \geq 4. \tag{2.218}$$

Doing so, comparison of (2.216)–(2.217) with (2.214)–(2.215) leads to

$$\tilde{\mathcal{B}} v = \rho_\infty b v - \sum_{i=1}^{3} (\rho_\infty - \rho_i) \langle bv, \phi_i \rangle_X \phi_i, \qquad \forall\, v \in \mathbb{C} \tag{2.219}$$

$$\tilde{\mathcal{C}} h = \rho_\infty^{-1} \mathcal{C} h - \sum_{i=1}^{3} (\rho_\infty^{-1} - \rho_i^{-1}) \langle c, \phi_i \rangle_X \langle h, \phi_i \rangle_X, \qquad \forall\, h \in X. \tag{2.220}$$

It was explained in Subsection 2.3.3 that the spillover can be divided into two different effects, namely the control spillover, for which the operator $\mathcal{B}_r K$ plays an essential role, and the observation spillover, that depends on $L\mathcal{C}_r$. Since these operators become changed by introducing the state transformation, such a change of the state allows to influence both kinds of spillover, which is the basic reason for the improvement of the perturbation estimate. This aspect will be discussed more in detail in Section 3.3.

Next, the compensator is designed on the basis of an approximation for $(\tilde{\mathcal{C}}, \mathcal{A}, \tilde{\mathcal{B}})$ such that the slowest eigenvalue $\lambda_1 = -\pi^2 - 1 = -10.87$ is shifted to $\lambda_{cl,1} = -13$ and all the other eigenvalues remain unaffected. In view of the decay of the eigenvalues (see (2.45)) the approximation model Σ_n (2.99)–(2.100) can be regarded to cover the dominant part of the system dynamics if it contains the first two modes, *i.e.*, $n = 2$. Applying (2.108)–(2.110) to $(\tilde{\mathcal{C}}, \mathcal{A}, \tilde{\mathcal{B}})$ yields

$$A_n = \operatorname{diag}(\lambda_1, \lambda_2) = \operatorname{diag}(-10.87, -13.87) \tag{2.221}$$

$$B_n = \begin{bmatrix} 0.824 \\ 0.060 \end{bmatrix} \tag{2.222}$$

$$C_n = \begin{bmatrix} 0.213 & 4.513 \end{bmatrix} \tag{2.223}$$

which gives a controllable and observable approximation Σ_n. The controller gain K is chosen as $K = [2.585 \ \ 0]$ so that it assigns the eigenvalues $\sigma(A_n - B_n K) = \{-13, \lambda_2\} = \{-13, -13.87\}$, and the observer gain $L = [20.260 \ \ -0.0109]^T$ provides $\sigma(A_n - LC_n) = \{-14, -15\}$. It has been found in Example 2.1-13 that $\mathcal{A}$ satisfies Assumption 2.1-9, which is required for applying Theorem 2.4-10, and that $\mathcal{A}$ is a sectorial Riesz-spectral operator. Therefore, also Assumption 2.3-2, which is the SDGA for the closed-loop system, holds. Finally, it has been shown in Example 2.1-13 that $\mathcal{A}$ is a Riesz-spectral operator. Taking this into account, Proposition 2.4-8 yields that Assumption 2.4-7

Table 2 – The unperturbed eigenvalues $\lambda_{cl,i} \in \sigma(A_n - B_n K) \cup \sigma(A_n - LC_n) \cup \sigma(\mathcal{A}_r)$ and closed-loop eigenvalues $\tilde{\lambda}_{cl,i} \in \sigma(\mathcal{A}_{cl})$ with largest real parts.

i	1	2	3	4	5	6	7
$\lambda_{cl,i}$	-13	-13.87	-14	-15	-18.87	-25.87	-40.48
$\tilde{\lambda}_{cl,i}$	-12.61	-13.87	-13.91	$-17.13 + j1.59$	$-17.13 - j1.59$	-25.87	-40.48

is satisfied. Thus, Theorem 2.4-10 can be applied. This shows that the closed-loop eigenvalues $\tilde{\lambda}_{cl,i}$ of $\mathcal{A}_{cl}$ lie entirely within the union of circles that are centered at the unperturbed eigenvalues $\lambda_{cl,i}$ of $\mathcal{A}_{cl,0}$ (see Table 2), and their radius $r = 5.11$ can be determined by (2.205) on the basis of a 60-th-order modal approximation. Due to their location in the complex plane as shown in Figure 13 it is assured that the closed-loop system has a stability margin of at least $\beta = 7.89$ and a damping of $d = 0.92$ in the worst case. The slowest 7 closed-loop eigenvalues $\tilde{\lambda}_{cl,i}$ are listed in Table 2 that were computed on the basis of a modal approximation of order 40. The actually occurring maximal difference to the desired values is $d = 2.36$ so that due to $r/d = 2.17$ the disk radius r is about 2 times larger than it would be in case of an ideal estimate. This shows that the proposed approach yields a viable eigenvalue estimate. Note, that the overestimation depends on the parameters ρ_i, so that these should be chosen with care.

◀

The system operator $\mathcal{A}_{cl}$ (see (2.175)) of the closed-loop dynamics Σ_{cl} has been divided in (2.177) into the sum of $\mathcal{A}_{cl,0}$ and Δ. This decomposition of the system operator is chosen with the intention that the estimate (2.203) is particularly sharp. For the further considerations in the next chapter, however, it is more convenient to use a different decomposition

$$\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta \tag{2.224}$$

with redefined operators

$$\mathcal{A}_{cl,0} := \begin{bmatrix} A_n - LC_n & 0 & 0 \\ 0 & A_n - B_n K & 0 \\ 0 & -\mathcal{B}_r K & \mathcal{A}_r \end{bmatrix}, \qquad \Delta := \begin{bmatrix} 0 & 0 & -L\mathcal{C}_r \\ B_n K & 0 & 0 \\ \mathcal{B}_r K & 0 & 0 \end{bmatrix}. \tag{2.225}$$

Theorem 2.4-10 remains valid for this new decomposition if the changed operators $\mathcal{A}_{cl,0}$ and Δ are used therein. The only difference is that the estimate for the spectrum

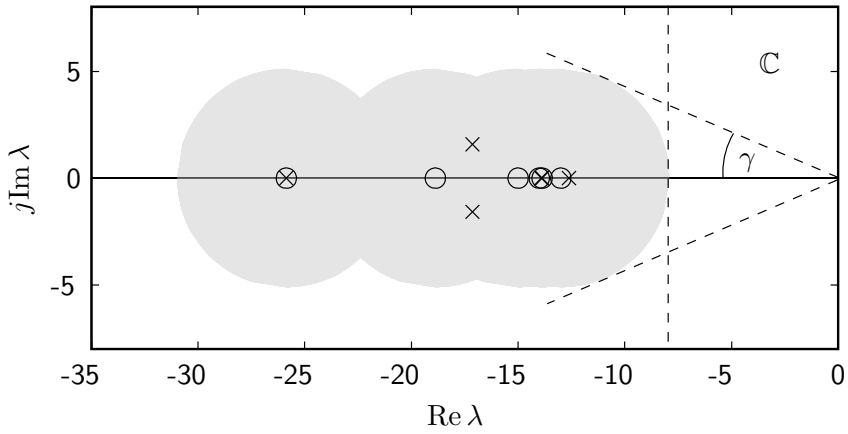

Figure 13 – Locations of the first six unperturbed eigenvalues $\lambda_{cl,i}$ ('$\circ$') and closed-loop eigenvalues $\tilde{\lambda}_{cl,i}$, $i = 1, 2, \ldots, 6$ ('$\times$') with largest real parts (compare to Table 2). The gray disks describe the estimated regions for the closed-loop eigenvalues that assure the stability margin $\beta = 7.89$ and the damping $d = \cos(\gamma) = 0.92$ due to $\gamma = 23.1°$.

perturbation is different. Instead of (2.203) one has now the more conservative upper bound

$$d \leq \tilde{d} \quad \text{with} \quad \tilde{d} := \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\|. \tag{2.226}$$

In the next chapter a control system design method is presented that allows to systematically reduce $\tilde{d}$.

Remark 2.4-14 In Subsection 2.3.4 it has been explained that asymptotic disturbance rejection can be assured by adding a signal model of the exogenous disturbance to the control loop. The analysis of the corresponding closed-loop spectrum can be done in an analog way as shown in this section. To this end, the operators in (2.225) have to be replaced by

$$\mathcal{A}_{cl,0} := \begin{bmatrix} A_n - LC_n & 0 & 0 & 0 \\ 0 & A_s & B_sC_n & 0 \\ 0 & -B_nK_s & A_n - B_nK & 0 \\ 0 & -\mathcal{B}_rK_s & -\mathcal{B}_rK & \mathcal{A}_r \end{bmatrix} \tag{2.227}$$

$$\Delta := \begin{bmatrix} 0 & 0 & 0 & -L\mathcal{C}_r \\ 0 & 0 & 0 & B_s\mathcal{C}_r \\ B_nK & 0 & 0 & 0 \\ \mathcal{B}_rK & 0 & 0 & 0 \end{bmatrix} \tag{2.228}$$

in order to describe the dynamics w.r.t. the state $x_{cl} := [e_n \;\; \tilde{x}_s \;\; x_n \;\; x_r]^T$, where $\tilde{x}_s$ is

the state of the signal model (see (2.184)). Note, that the sub-block

$$\tilde{A}_n = \begin{bmatrix} A_s & B_s C_n \\ -B_n K_s & A_n - B_n K \end{bmatrix} \tag{2.229}$$

in $\mathcal{A}_{cl,0}$ has the eigenvalues that are assigned by the compensator to the extended approximation (2.186)–(2.188). Thus, $\mathcal{A}_{cl,0}$ has the desired eigenvalues due to its block-diagonal structure. Using these modified operators the statements of this section as well as those of the next chapter remain valid. ◀

Chapter 3

Spillover reduction for continuous-time control

The detrimental influence of the residual dynamics on the closed-loop behavior has the consequence that an iterative design procedure may be needed: After an initial compensator design, the analysis of the closed-loop system may yield that its behavior is unsatisfying so that the design has to be repeated with modified parameters, until the closed-loop behavior meets the specifications. Since it is not clear a priori in which way the redesign has to be modified this approach can hardly be considered as systematic. This motivates to seek for an approach that reduces the spillover effectively and systematically (see [2, 105]).

A quite natural approach is to increase the order of the approximation so that less of the system dynamics is neglected, so that in consequence the residual dynamics has less influence on the closed-loop behavior. Alternatively, the same can be achieved by reconstructing some of the residual modal states by means of a so-called *residual mode filter*, that acts as an observer for single eigenmodes without correction term. These reconstructed states can be used to eliminate their contributions to the measurement y (see [13]). Consequently, those modes do then not cause observation spillover since they do not excite the observation error (see (2.170)). In principle, the spillover can be made arbitrarily small in these ways by increasing the order of the compensator or the residual mode filter. However, for these approaches no a priori estimates are available that yield the required order for assuring prescribed performance criteria so that the design procedure still has to be done in a "trial and error" manner. Furthermore, it turns out that typically a comparatively large number of residual modes have to be

eliminated and thus a high order of the compensator or the filter may be needed for a satisfying reduction of the spillover (see Example 2.1-15). In case that it is desired that the total order of the compensator is low, a more efficient approach for spillover reduction is thus needed.

In this chapter the classical early-lumping design scheme for observer-based continuous-time control is improved by combining it with a new effective spillover reduction technique. The reduction relies basically on the results in [56, 77], where it is shown that certain linear combinations of the modal states can be reconstructed by means of a finite-dimensional dynamical system, which is named *output observer*. These linear combinations have the essential characteristic that the error of the reconstruction can be assured to decay exponentially. This makes it possible to use them as *fictitious outputs*, because the error dynamics is separated from the remaining closed-loop dynamics so that both dynamics can be designed independently from each other. It turns out that the fictitious outputs contain less contributions of the residual modes than the measurable output so that the spillover becomes reduced when they are used for an observer-based compensator (see Figure 14a). In this chapter the results of [56] are extended in two directions. First, it is shown that the spillover reducing effect can be increased by using several output observers that are arranged in a cascaded structure (see Figure 14b). Second, the properties of the resulting closed-loop spectrum are analyzed using the results of Section 2.4.

A comparison of the spectrum perturbations of the closed-loop with and without an output observer yields a measure for the achieved spillover suppression. It turns out that an upper bound for the spectrum perturbation decreases exponentially with respect to the order of the additional dynamics that is used for the reconstruction of the fictitious output. For the classical spillover reduction approaches by increasing the order of the approximation or the residual mode filter, in contrast, it is not possible to describe the spectrum perturbation by a simple relation. Moreover, the proposed method proves to be significantly more effective in many applications so that the approach is useful when low order controllers are desired. Furthermore, an a priori upper bound for the total compensator order, that assures a prescribed reduction of the spectrum perturbation, is given.

Inserting output observers in the control loop has the effect that both the observation and the control spillover can be reduced. It is shown that the impact on these two kinds of spillover depends on the choice of the state variables of the plant. However,

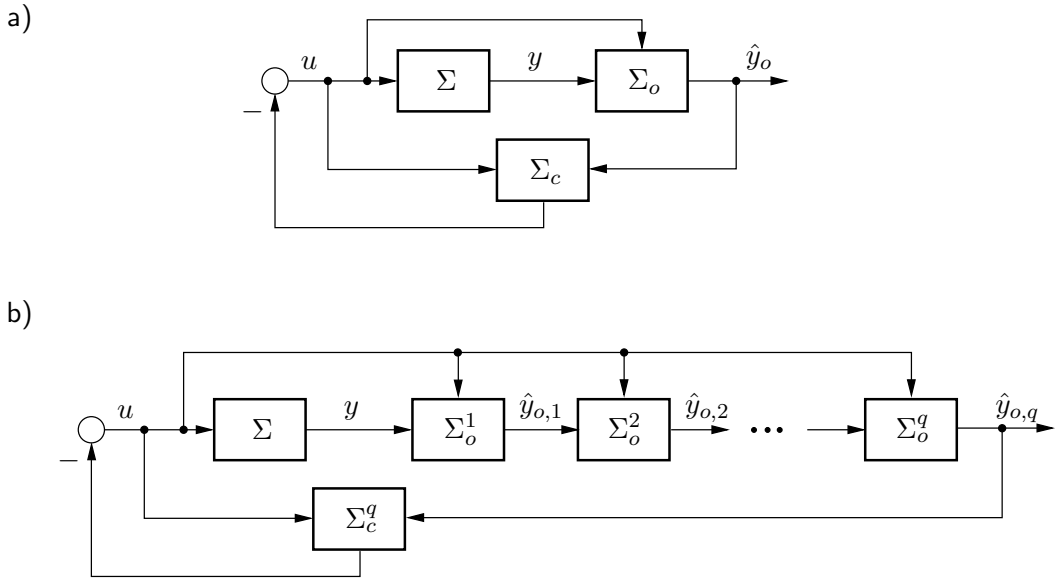

Figure 14 – Structure of the closed-loop system with the infinite-dimensional system Σ and with a) a single output observer Σ_o and b) several cascaded output observers $\Sigma_o^1, \ldots, \Sigma_o^q$. The observer-based compensators Σ_c and Σ_c^q use the resulting reconstructions $\hat{y}_o$ and $\hat{y}_{o,q}$, respectively.

the resulting reduction of the spectrum perturbation is the same, regardless if the approach is applied for the suppression of observation or control spillover.

In the first section of this chapter the reconstruction of fictitious outputs is introduced. The design of the compensator on the basis of a system approximation for the fictitious output and the analysis of the resulting spillover suppression are subject of Section 3.2, where also a cascaded arrangement of output observers is considered. The approach is presented in this section such that solely the observation spillover becomes reduced. It is shown in Section 3.3 how the approach can be used also for the reduction of control spillover.

3.1 Reconstruction of fictitious outputs

In Section 2.3.3 it has been discussed how the residual dynamics, that is not contained in the approximation and thus neglected during the compensator design, influences the spectrum $\sigma(\mathcal{A}_{cl})$ of the closed-loop system operator, and an estimate for the spectrum perturbation was given in Section 2.4.2. The causation for the spillover effect becomes

apparent if the dynamics (2.170) of the observation error $e_n(t) = x_n(t) - \hat{x}_n(t)$ are considered. It shows that the output $y_r(t) = \mathcal{C}_r x_r(t)$ of the residual dynamics excites the error dynamics so that the observer cannot converge correctly because the residual dynamics itself is excited by the observation error due to the control law $u(t) = -K\hat{x}_n(t) = -Kx_n(t) + Ke_n(t)$ (see (2.169)).

In order to reduce the spillover it is for this reason the core idea to design the observer on the basis of a fictitious output y_o instead of the measurable output y, where the contributions of the residual state to y_o are reduced compared to y. The spillover can then expected to be small since the observer is less affected. In terms of the closed-loop dynamics (2.174)–(2.177) this can be understood when it is taken into account that suppressing the contributions of the residual state corresponds to reducing the norm of $\mathcal{C}_r$ in (2.177) so that the spectrum $\sigma(\mathcal{A}_{cl,0})$, that contains the desired closed-loop eigenvalues, is less perturbed (see Theorem 2.4-10).

If the error dynamics of the output observer can be excited by the input u or the states x and $\hat{x}$, the usage of the output observer will cause additional spillover and thus might destabilize the control loop. It is therefore essential for the approach that the error dynamics are homogeneous and asymptotically stable so that the error decays and cannot be excited. In this case the *separation principle* holds for the output observer which means that the spectra of the output observer dynamics and the dynamics of the remaining closed-loop system are independent from each other. Hence, the output observer does then not cause additional spillover in the closed-loop system.

Let the fictitious output operator that corresponds to y_o be denoted as $\mathcal{C}_1 : X \mapsto \mathbb{C}^{n_o}$, *i.e.*,

$$y_o(t) = \mathcal{C}_1 x(t), \qquad t \geq 0. \tag{3.1}$$

Since the fictitious output y_o is not available by measurement, it must be possible to generate this output by means of a finite-dimensional dynamic system

$$\Sigma_o : \quad \dot{\hat{y}}_o(t) = A_o\hat{y}_o(t) + B_o u(t) + L_o y(t), \qquad t > 0, \quad \hat{y}_o(0) = \hat{y}_{o,0} \in \mathbb{C}^{n_o}. \tag{3.2}$$

Such a system is called *output observer* Σ_o (see [56, 77]). It provides the reconstructed output $\hat{y}_o$ that converges toward the fictitious output y_o. In Section 3.2 a compensator will be designed that uses $\hat{y}_o$ instead of y in which way the spillover turns out to be reduced (see Figure 14a). Due to the above discussion the output operator $\mathcal{C}_1$ in (3.1) must satisfy the following two requirements:

1. The output y_o can be reconstructed asymptotically, *i.e.*, $\hat{y}_o(t) \to y_o(t)$, by means of a finite-dimensional output observer Σ_o according to (3.2).
2. The contribution of x_r to y_o must be suppressed.

In the following subsection a class of fictitious outputs, that satisfy Item 1, is identified. Item 2 is addressed afterwards in Subsection 3.1.2.

3.1.1 Characterization of reconstructible fictitious outputs

While an infinite-dimensional Luenberger observer for the plant would reconstruct the whole state x so that an output y_o for any fictitious output operator $\mathcal{C}_1$ can be obtained, one cannot expect that this is possible with an output observer of the form (3.2) with finite order n_o. For that reason it will be analyzed under which conditions on $\mathcal{C}_1$ it is possible to satisfy Item 1. More precisely, the following problem is considered.

Problem 3.1-1

Find a bounded linear operator $\mathcal{C}_1 : X \mapsto \mathbb{C}^{n_o}$ such that there exist matrices A_o, B_o, and L_o for which the output observer Σ_o in (3.2) reconstructs $y_o(t) = \mathcal{C}_1 x(t)$, *i.e.*,

$$\lim_{t\to\infty} \hat{y}_o(t) = y_o(t), \qquad \forall\, y_o(0), \hat{y}_o(0) \in \mathbb{C}^{n_o} \tag{3.3}$$

holds. ◀

In order to solve this problem, $A_o \in \mathbb{C}^{n_o \times n_o}$, $B_o \in \mathbb{C}^{n_o \times p}$, and $L_o \in \mathbb{C}^{n_o \times m}$ are designed such that the dynamics of the error

$$e_o(t) := y_o(t) - \hat{y}_o(t), \qquad t \geq 0 \tag{3.4}$$

are homogeneous and exponentially stable. These dynamics are obtained from (3.2), $\dot{y}_o(t) = \mathcal{C}_1 \dot{x}(t) = \mathcal{C}_1 \mathcal{A} x(t) + \mathcal{C}_1 \mathcal{B} u(t)$, and $y(t) = \mathcal{C} x(t)$, yielding

$$\dot{e}_o(t) = -A_o \hat{y}_o(t) + (\mathcal{C}_1 \mathcal{A} - L_o \mathcal{C}) x(t) + (\mathcal{C}_1 \mathcal{B} - B_o) u(t), \tag{3.5}$$

and insertion of

$$-\hat{y}_o(t) = e_o(t) - y_o(t) = e_o(t) - \mathcal{C}_1 x(t) \tag{3.6}$$

gives

$$\dot{e}_o(t) = A_o e_o(t) + (\mathcal{C}_1 \mathcal{A} - L_o \mathcal{C} - A_o \mathcal{C}_1) x(t) + (\mathcal{C}_1 \mathcal{B} - B_o) u(t). \tag{3.7}$$

Condition (3.3) implies that neither x nor u can excite the error e_o. Therefore, it is apparent from (3.7) that

$$B_o v \stackrel{!}{=} \mathcal{C}_1 \mathcal{B} v, \qquad \forall\, v \in \mathbb{C}^p \tag{3.8}$$

$$\mathcal{C}_1 \mathcal{A} h - L_o \mathcal{C} h - A_o \mathcal{C}_1 h \stackrel{!}{=} 0, \qquad \forall\, h \in D(\mathcal{A}) \tag{3.9}$$

has to be satisfied, in which case (3.7) simplifies to

$$\dot{e}_o(t) = A_o e_o(t), \qquad t > 0. \tag{3.10}$$

Thus, A_o has to be a Hurwitz matrix[1], so that the error e_o decays and consequently (3.3) holds. These considerations yield the following general result (see also [56, 99]).

Lemma 3.1-2

Let A_o be a Hurwitz matrix and suppose that L_o and $\mathcal{C}_1$ are such that (3.9) holds. Then, $\mathcal{C}_1$ solves Problem 3.1-1 and the corresponding B_o is given by (3.8).

The *Sylvester-operator equation* (3.9) has a unique solution if and only if the spectra of $\mathcal{A}$ and A_o are disjoint (see [100]). Furthermore, in [56] a parametric solution of (3.9) was proposed. However, the special choice

$$A_o = \mu I, \qquad \mu < 0 \tag{3.11}$$

will be used in the following to simplify the derivation of the results, in which I denotes the identity matrix on $\mathbb{C}^{n_o \times n_o}$. Then, (3.9) can be solved for $\mathcal{C}_1$ yielding

$$\mathcal{C}_1 h = L_o \mathcal{C} (\mathcal{A} - \mu I)^{-1} h, \qquad \forall\, h \in X, \tag{3.12}$$

where $L_o \in \mathbb{C}^{n_o \times m}$ can be chosen arbitrarily and μ has to satisfy

$$\mu \in S := (-\infty, 0) \cap \rho(\mathcal{A}) \tag{3.13}$$

for ensuring that the resolvent in (3.12) exists and that A_o is a real Hurwitz matrix (for the definition of $\rho(\mathcal{A})$ see Subsection 2.1.3). The corresponding B_o according to (3.8) becomes

$$B_o = L_o \mathcal{C} (\mathcal{A} - \mu I)^{-1} \mathcal{B} = -L_o G(\mu) \tag{3.14}$$

[1] A matrix M is called a *Hurwitz matrix* if $\operatorname{Re} \lambda < 0$, $\forall\, \lambda \in \sigma(M)$, holds.

with

$$G(s) = \mathcal{C}(sI - \mathcal{A})^{-1}\mathcal{B} \tag{3.15}$$

denoting the transfer matrix of the system Σ (see, *e.g.*, [43]). In conclusion, the following result has been obtained.

Theorem 3.1-3

Suppose that $A_o = \mu I$ and $B_o = -L_o G(\mu)$ for arbitrary $L_o \in \mathbb{C}^{n_o \times m}$ and $\mu \in S$ (see (3.13)). Then, the output observer Σ_o in (3.2) reconstructs the output

$$y_o(t) = \mathcal{C}_1 x(t) = L_o \mathcal{C}(\mathcal{A} - \mu I)^{-1} x(t), \qquad t \geq 0 \tag{3.16}$$

asymptotically, at which

$$\dot{e}_o(t) = \mu e_o(t), \qquad t > 0 \tag{3.17}$$

holds for the error $e_o(t) = y_o(t) - \hat{y}_o(t)$, $t \geq 0$.

An important question in this regard is whether it is possible to reconstruct outputs that contain only the contribution of a single modal state. If this is possible, these outputs can be used directly for the implementation of state feedback control resulting from the early-lumping approach. In this way the spillover could be avoided completely. In order to analyze that question, the case $m = n_o = 1$ is considered for simplicity, and $\mathcal{A}$ is assumed to be a Riesz-spectral operator. In order to reconstruct the modal state x_i^* by the output observer, *i.e.*, $x_i^*(t) = y_o(t) = \mathcal{C}_1 x(t)$, relation $x_i^*(t) = \langle x(t), \psi_i \rangle_X$, $i \in \mathbb{N}$, (see (2.102)) is used that makes $\mathcal{C}_1 h = \langle h, \psi_i \rangle_X$ apparent. Inserting this output operator and $A_o = \mu I$ into the Sylvester-operator equation (3.9) yields

$$\langle \mathcal{A}h, \psi_i \rangle_X - L_o \mathcal{C}h - \mu \langle h, \psi_i \rangle_X = 0, \qquad \forall\, h \in D(\mathcal{A}). \tag{3.18}$$

By taking

$$\langle \mathcal{A}h, \psi_i \rangle_X = \langle h, \mathcal{A}^* \psi_i \rangle_X = \lambda_i \langle h, \psi_i \rangle_X \tag{3.19}$$

into account, for which it is used that ψ_i is an eigenvector of $\mathcal{A}^*$ corresponding to the eigenvalue $\overline{\lambda_i}$ (see Footnote 6 on page 16), (3.18) can be rearranged as

$$(\lambda_i - \mu) \langle h, \psi_i \rangle_X = L_o \mathcal{C}h, \qquad \forall\, h \in D(\mathcal{A}). \tag{3.20}$$

This finally leads to

$$x_i^*(t) \stackrel{!}{=} \frac{L_o}{\lambda_i - \mu} y(t) \tag{3.21}$$

when h is replaced by $x(t)$. This shows that y_o coincides with the modal state x_i^* only if the system output y is proportional to the same modal state. In this case, however, which is very restrictive for the applications, no spillover would be present anyway when y is used as compensator input so that the output observer does not gain any advantage. Thus, except for this degenerate case the reconstruction of single modal states is not possible by means of the output observer Σ_o. Nevertheless, the reconstructed output y_o has the property to suppress the contributions of the residual modal states in many cases as was claimed before. This is discussed next.

3.1.2 Suppression of the residual modal state contributions

It has been argued at the beginning of Section 3.1 that the spillover can be expected to be reduced by using a fictitious output y_o for the compensator design if the contributions of the residual modal states to y_o are suppressed compared to the contributions to y. This reduction of the contributions is analyzed now, at which its extent depends on the application so that only a qualitative discussion is possible. The situation is most transparent if $\mathcal{A}$ is a Riesz-spectral operator. For more general systems the spillover reducing effect of the output observer becomes apparent in Subsection 3.2.2. If $\mathcal{A}$ is a Riesz-spectral operator, it has the representation

$$\mathcal{A}h = \sum_{i=1}^{\infty} \lambda_i \langle h, \psi_i \rangle_X \phi_i, \qquad \forall\, h \in D(\mathcal{A}) \tag{3.22}$$

(see (2.33)) with ϕ_i and ψ_i denoting the eigenvectors of $\mathcal{A}$ and $\mathcal{A}^*$, respectively. Then, c_i in (2.15) and x can be written as

$$c_i = \sum_{j=1}^{\infty} c_{i,j}^* \psi_j, \qquad i = 1, 2, \ldots, m, \quad c_{i,j}^* \in \mathbb{C} \tag{3.23}$$

$$x(t) = \sum_{i=1}^{\infty} x_i^*(t) \phi_i, \qquad t \geq 0, \quad x_i^*(t) \in \mathbb{C}, \tag{3.24}$$

wherein x_i^* are the modal states (see (2.22) and (2.102)). Using this, one obtains for the k-th output

$$y_k(t) = \langle x(t), c_k \rangle_X = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i^*(t) \overline{c_{k,j}^*} \langle \phi_i, \psi_j \rangle_X = \sum_{i=1}^{\infty} \overline{c_{k,i}^*}\, x_i^*(t), \tag{3.25}$$

wherein the biorthonormality of ϕ_i and ψ_i (see (2.21)) has been used. In contrast, when $(\mathcal{A} - \mu I)^{-1} h = \sum_{i=1}^{\infty} \frac{1}{\lambda_i - \mu} \langle h, \psi_i \rangle_X \phi_i,\ \forall\, h \in X$, following from (3.22), is taken

into account, the k-th fictitious output $y_{o,k}$ is obtained from (3.16), (2.15), and (3.23)–(3.24) yielding

$$y_{o,k}(t) = \sum_{i=1}^{\infty} \frac{\overline{c^*_{k,i}}}{\lambda_i - \mu}\, x^*_i(t), \tag{3.26}$$

where the case $n_o = m$ and $L_o = I$ has been assumed for simplicity. In the more general case $n_o \neq m$ and $L_o \neq I$ one has to replace $c^*_{k,i}$ in (3.26) by

$$\tilde{c}^*_{k,i} = l^T_{o,k} \begin{bmatrix} c^*_{1,i} \\ \vdots \\ c^*_{m,i} \end{bmatrix} \tag{3.27}$$

with $l^T_{o,k}$ denoting the k-th row of $\overline{L_o}$. Comparison of (3.25) and (3.26) shows that the absolute values $|x^*_i \overline{c^*_{k,i}}|$ of the contributions of the modal states x^*_i are reduced in the output $y_{o,k}$ by $1/|\lambda_i - \mu|$. This factor decreases fast in many applications for increasing index i so that the modes of the residual dynamics Σ_r do less contribute to y_o than to y. Note, that it is not required for the spillover reducing effect that the real parts of the eigenvalues of $\mathcal{A}$ decay, but instead, their absolute values have to increase. For instance, the eigenvalues λ_i of the Euler-Bernoulli beam in Example 2.1-14 and those of the heat conductor in Example 2.2-1 satisfy $|\lambda_i| \sim i^2$ so that $1/|\lambda_i - \mu| \to 0$ for $i \to \infty$. For that reason, the usage of the fictitious output y_o for feedback control can be expected to cause less spillover than the use of the output y. This is the basic concept of the spillover suppressing compensator design presented in [56]. This approach is further refined in the next section to obtain a systematic spillover reduction scheme.

In the following example the reconstruction of a fictitious output for the Euler-Bernoulli beam with Kelvin-Voigt damping considered in Example 2.1-15 is demonstrated. Since this system has an accumulation point in its spectrum (see Figure 6), the contributions $x^*_i(t)\overline{c^*_{k,i}}/(\lambda_i - \mu)$ to the fictitious output y_o do not become arbitrarily small for $i \to \infty$, which is why the accumulation point reduces the effectiveness of the approach. It will be demonstrated, that the spillover is reduced significantly, nevertheless.

Example 3.1-4 (Euler-Bernoulli beam with Kelvin-Voigt damping, continued)

For the Euler-Bernoulli beam with Kelvin-Voigt damping, that was considered in Example 2.1-15, a fictitious output shall be reconstructed by aid of an output observer Σ_o with order $n_o = 1$. For the length $\ell = 1$ and the damping constant $\delta = 0.004$ the most dominant eigenvalues of the system operator $\mathcal{A}$ are $\lambda_{\pm 1} = -0.39 \pm j9.86$, $\lambda_{\pm 2} = -6.23 \pm j38.98$, $\lambda_{\pm 3} = -31.56 \pm j83.03$ according to (2.68). Thus, choosing

$A_o = \mu = -4$, the observer dynamics is sufficiently fast compared to the system dynamics in view of (3.17). Since the observer gain $L_o \in \mathbb{C}^{n_o \times m}$ may be chosen arbitrarily, $L_o = 1$ is a natural choice due to $n_o = m = 1$. Thus, it remains to determine

$$B_o = -G(\mu) = -\mathcal{C}(\mu I - \mathcal{A})^{-1}\mathcal{B}. \tag{3.28}$$

Therein, $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are given by (2.66), (2.56), and (2.58), respectively, in which the input and output distribution functions b and c are given by

$$b(z) = \frac{1}{\beta_2 - \beta_1} \cdot \mathbf{1}_{[\beta_1,\beta_2]}(z), \qquad c(z) = \frac{1}{\gamma_2 - \gamma_1} \cdot \mathbf{1}_{[\gamma_1,\gamma_2]}(z) \tag{3.29}$$

with $\beta_1 = 0.25 - 10^{-3}$, $\beta_2 = 0.25 + 10^{-3}$, $\gamma_1 = 0.75 - 10^{-3}$, and $\gamma_2 = 0.75 + 10^{-3}$. Instead of computing the inverse in (3.28) the equation

$$(\mu I - \mathcal{A})h = \begin{bmatrix} \mu & -\mathcal{A}_0 \\ \mathcal{A}_0 & \mu + 2\delta\mathcal{A}_0^2 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \stackrel{!}{=} \begin{bmatrix} 0 \\ b \end{bmatrix} \tag{3.30}$$

is solved, which yields $h = (\mu I - \mathcal{A})^{-1}\mathcal{B}$, and then

$$B_o = -G(\mu) = -\mathcal{C}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = -\left\langle \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \begin{bmatrix} \mathcal{A}_0^{-1}c \\ 0 \end{bmatrix} \right\rangle_X \tag{3.31}$$

is evaluated (see (2.58)). Assuming a beam length $\ell = 1$ and taking $\mathcal{A}_0 h_i = h_i''$, $h_i \in D(\mathcal{A}_0)$, $i = 1, 2$, with $D(\mathcal{A}_0)$ defined in (2.53) into account, (3.30) amounts to solve the boundary value problem

$$\mu h_1(z) - h_2''(z) = 0 \tag{3.32}$$

$$h_1''(z) + \mu h_2(z) + 2\delta h_2''''(z) = b(z) \tag{3.33}$$

$$h_1(0) = h_1(1) = h_2''(0) = h_2''(1) = 0, \tag{3.34}$$

which can be done by means of symbolic computing software. In a second step, B_o is obtained from (3.31) which becomes

$$B_o = -\langle h_1, \mathcal{A}_0^{-1}c\rangle_{L_2} = -\langle \mathcal{A}_0^{-1}h_1, c\rangle_{L_2} \tag{3.35}$$

by the definition of the inner product (2.52) and the self-adjointness of $\mathcal{A}_0$. Using $\mathcal{A}_0^{-1}h_1 = \frac{1}{\mu}h_2$, following from (3.32), and $\langle g, h\rangle_{L_2} = \int_0^\ell g(z)\overline{h(z)}\mathrm{d}z$ yields

$$B_o = -\frac{1}{\mu}\langle h_2, c\rangle_{L_2} = -\frac{500}{-4}\int_{\gamma_1}^{\gamma_2} h_2(z)\mathrm{d}z = -0.006275, \tag{3.36}$$

where (3.29) has been taken into account. In order to show that the fictitious output $y_o(t) = L_o\mathcal{C}(\mathcal{A} - \mu I)^{-1}x(t) = \mathcal{C}(\mathcal{A} + 4I)^{-1}x(t)$ (see (3.16)), that is reconstructed by

Table 3 – Comparison of the first ten modal state weights $|c_i^*|$ of the measurable output y and the modal state weights $|c_{o,i}^*|$ of the fictitious output y_o.

i	1	-1	2	-2	3	-3	4	-4	5	-5
$\lvert c_i^*\rvert/\lvert c_1^*\rvert$	1	0.101	0.354	0.0090	0.111	0.0013	0	0	0.040	$1.62 \cdot 10^{-4}$
$\lvert c_{o,i}^*\rvert/\lvert c_{o,1}^*\rvert$	1	0.101	0.095	0.0024	0.013	0.0002	0	0	0.002	$0.07 \cdot 10^{-4}$

the designed output observer, suppresses the higher modal states, the absolute values of the modal state weights c_i^* and $c_{o,i}^* := c_i^*/(\lambda_i - \mu)$ (see (3.25)–(3.26)) are given in Table 3 for $i = \pm 1, \pm 2, \ldots, \pm 5$, scaled by the first weights $|c_1^*|$ and $|c_{o,1}^*|$, respectively. The comparison of these coefficients shows that the high frequency modal states do significantly less contribute to the fictitious output y_o than to the measurable output y. The same result is obtained when the step responses with respect to y and y_o are compared to the step responses of modal approximations with respect to y and y_o. In Figure 15 these are plotted for modal approximations with order $n = 2$. In the subfigure (a) the output y and the output y_n of the approximation differ significantly. This error corresponds to the contributions of the modal states $x_{\pm i}^*(t)$, $i > 2$. As shown in the subfigure (b) the difference and thus the contribution of the residual dynamics is much smaller when the fictitious output y_o and the corresponding approximation is used. Therefore, one can expect that an observer-based compensator designed on the basis of y_o is less affected by the residual dynamics so that the spillover effect becomes small. This is the basic idea of the design procedure presented in the next section. ◀

3.1.3 Extended System with the reconstructed output

In the following, it is investigated how the reconstruction of a fictitious output influences the controller design. To this end, consider the *extended system* Σ_e consisting of the plant Σ and the output observer Σ_o (see Figure 16). Its state-space model with the states x and e_o reads

$$\Sigma_e: \quad \dot{x}(t) = \mathcal{A}x(t) + \mathcal{B}u(t), \qquad t > 0, \quad x(0) = x_0 \in X \tag{3.37}$$

$$\dot{e}_o(t) = \mu\, e_o(t), \qquad t > 0, \quad e_o(0) = e_{o,0} \in \mathbb{C}^{n_o} \tag{3.38}$$

$$\hat{y}_o(t) = \mathcal{C}_1 x(t) - e_o(t), \qquad t \geq 0 \tag{3.39}$$

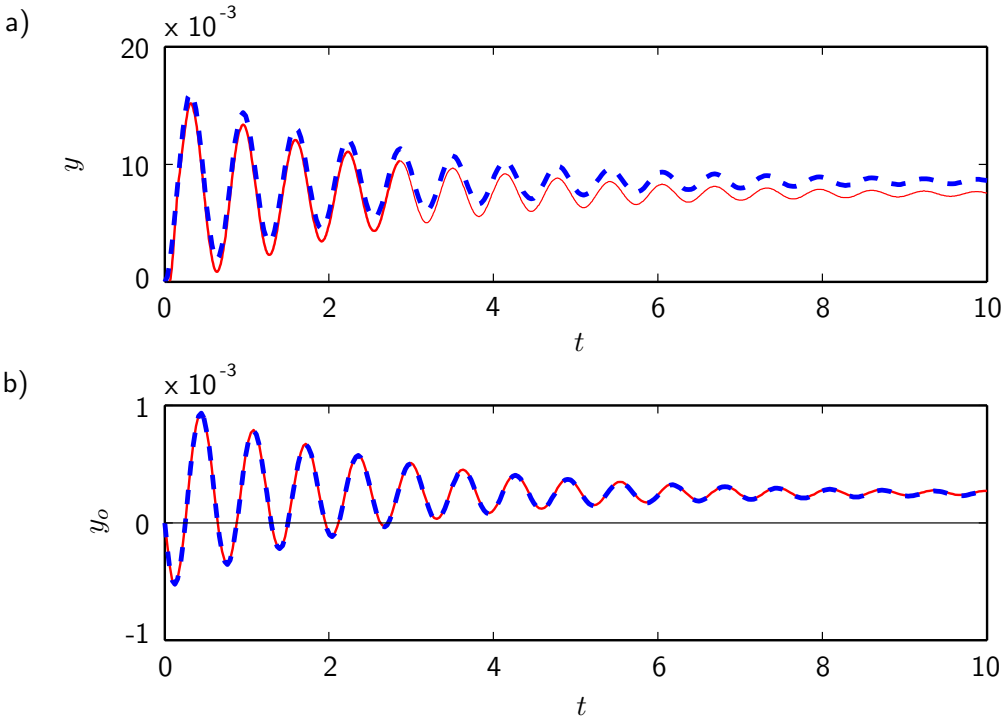

Figure 15 – Step responses of the Euler-Bernoulli beam with Kelvin-Voigt damping with respect to (a) the measurable output y and (b) the fictitious output y_o. The solid lines show the behavior of reference modal approximations of order $n_{high} = 60$ and the dashed lines correspond to modal approximations with low order $n = 2$.

with

$$\mathcal{C}_1 = L_o\mathcal{C}(\mathcal{A} - \mu I)^{-1} \tag{3.40}$$

(see (2.3), (3.1), and (3.16)–(3.17)). Due to the fact that the dynamics of e_o are homogeneous, they are uncontrollable and do therefore not contribute to the transfer behavior of the extended system Σ_e. Since only the transfer behavior is relevant for the compensator design while uncontrollable subsystems do not affect the impact of the controller, it is essential for the design approach proposed in Section 3.2 to observe from (3.37)–(3.39) that the transfer behavior of Σ_e equals that of Σ when $\mathcal{C}$ in (2.4) is replaced by the output operator $\mathcal{C}_1$.

3.2 Compensator design using output observers

In this section an observer-based compensator will be designed using a fictitious output that is reconstructed by an output observer. The corresponding structure of the closed-loop system is depicted in Figure 16. It has been shown in the previous section that the output observer suppresses the contributions of the residual modes to the reconstructed fictitious. This fact will be exploited in the following to reduce the spillover, where the extent of the reduction is analyzed in terms of the spectrum perturbation. The basic idea to reduce the spillover by adding an output observer to the closed-loop can be used in a repeated manner. Thus, after the spillover reduction technique has been presented for the use of a single output observer, the approach is extended by utilizing several output observers. This ends up in a systematic approach to assure a predefined spectrum perturbation that can be summarized as follows: The observer-based compensator, that was determined in Subsection 2.3.2, forms the starting point of the design approach. The resulting spillover will then be suppressed by adding a sufficient number of output observers into the control loop. Since the compensator is fed by a reconstructed output instead of the plant output, the observer gain has to be adapted such that the eigenvalues of the controlled approximation, remain the same that were used for the compensator design. Thus, the spillover reduction technique consists of two parts: first, adding successively output observers until the spectrum perturbation becomes sufficiently small, and second, adapting the initially designed compensator to the fictitious output.

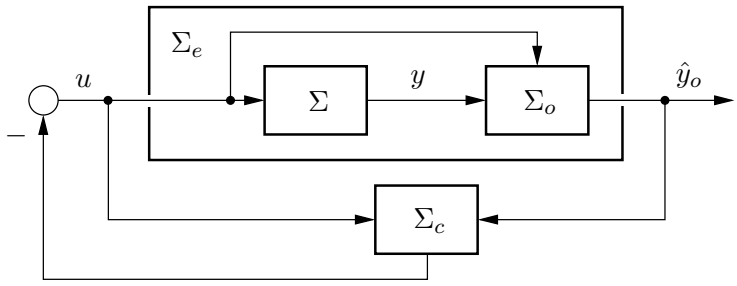

Figure 16 – Structure of the closed-loop system with an output observer Σ_o whose output $\hat{y}_o$ is used by the observer-based compensator Σ_c.

3.2.1 Observer-based state feedback control using a single output observer

In order to simplify the following considerations the gain L_o and the dynamic matrix A_o of the output observer and its order n_o are chosen as $L_o = I$, $A_o = \mu I$, and $n_o = m$, wherein m is the number of outputs of system Σ. Thus, the output observer (3.2) has the form

$$\Sigma_o: \quad \dot{\hat{y}}_o(t) = \mu \hat{y}_o(t) + B_o u(t) + y(t), \qquad t > 0, \quad \hat{y}_o(0) = \hat{y}_{o,0} \in \mathbb{C}^m, \tag{3.41}$$

and the output operator of the extended system Σ_e becomes $\mathcal{C}_1 = \mathcal{C}(\mathcal{A} - \mu I)^{-1}$ (see (3.37)–(3.40)). The compensator is designed on the basis of an n-dimensional approximation of Σ_e. Note, that the subsystem of Σ_e that corresponds to the observation error e_o is uncontrollable (see (3.38)) so that it does not have to be taken into account for the approximation since only the transfer behavior plays a role for the control performance (see Section 3.1.3). A modal approximation of Σ_e with $e_o \equiv 0$ is obtained from (2.108)–(2.110) when the output operator $\mathcal{C}$ therein is replaced by $\mathcal{C}_1 = \mathcal{C}(\mathcal{A} - \mu I)^{-1}$. It is straightforward to verify that doing so yields the approximation

$$\Sigma_n^1: \quad \dot{x}_n(t) = A_n x_n(t) + B_n u(t), \qquad t > 0,\ x_n(0) = \mathcal{F}^{-1}\mathcal{P}x_0 \in \mathbb{C}^n \tag{3.42}$$

$$y_{n,1}(t) = C_{n,1} x_n(t), \qquad t \geq 0 \tag{3.43}$$

with

$$C_{n,1} = C_n (A_n - \mu I)^{-1}. \tag{3.44}$$

For the analysis of the closed-loop behavior a state space representation of the residual dynamics Σ_r^1 of the extended system is needed that describes the error of the approximation Σ_n^1. That means that the output $y_{n,1}$ of Σ_n^1 and the output $y_{r,1}$ of Σ_r^1 are complementary w.r.t. the output $\hat{y}_o$ of Σ_e, *i.e.*,

$$y_{n,1}(t) + y_{r,1}(t) = \hat{y}_o(t), \qquad \forall\, t \geq 0. \tag{3.45}$$

In Appendix A.7 it is shown that the following state space model describes the residual dynamics Σ_r^1.

Lemma 3.2-1

The infinite-dimensional system

$$\Sigma_r^1: \quad \dot{x}_r(t) = \mathcal{A}_r x_r(t) + \mathcal{B}_r u(t), \qquad t > 0,\ x_r(0) = (I - \mathcal{P})x_0 \in X_r \tag{3.46}$$

$$\dot{e}_o(t) = \mu e_o(t), \qquad t > 0,\ e_o(0) = e_{o,0} \in \mathbb{C}^m \tag{3.47}$$

$$y_{r,1}(t) = \mathcal{C}_{r,1} x_r(t) - e_o(t), \qquad t \geq 0 \tag{3.48}$$

with

$$\mathcal{C}_{r,1} = \mathcal{C}_r(\mathcal{A}_r - \mu I)^{-1} \tag{3.49}$$

satisfies (3.45) and thus describes the residual dynamics of the extended system Σ_e.

Comparison of Σ_n^1 with the approximation Σ_n of the infinite-dimensional plant Σ shows that the matrices A_n and B_n are the same but the output matrix C_n is different (see (2.99)–(2.100)). Therefore, Σ_n^1 is stabilizable, respectively controllable, if Σ_n is. In contrast, it is less obvious whether $(C_{n,1}, A_n)$ is observable. The following lemma gives an answer to this question.

Lemma 3.2-2

Suppose $\mu \in S$ *with* S *according to (3.13). Then,* $(C_{n,1}, A_n)$ *is detectable if and only if* (C_n, A_n) *is detectable. Furthermore,* $(C_{n,1}, A_n)$ *is observable if and only if* (C_n, A_n) *is observable.*

The proof is given in Appendix A.8. Since observability of (C_n, A_n) was assumed before so that also $(C_{n,1}, A_n)$ is observable, an observer-based compensator

$$\Sigma_c^1: \quad \dot{\hat{x}}_n(t) = (A_n - L_1 C_{n,1})\hat{x}_n(t) + B_n u(t) + L_1 \hat{y}_o(t), \quad t > 0,\ \hat{x}_n(0) \in \mathbb{C}^n \tag{3.50}$$

$$u(t) = -K\hat{x}_n(t), \qquad t \geq 0 \tag{3.51}$$

may be designed for Σ_n^1 that uses the output $\hat{y}_o$ provided by the output observer Σ_o. For the assignment of the desired eigenvalues of $A_n - B_n K$ the state feedback gain K remains the same as for the closed-loop system without output observer since A_n and B_n have not been changed. In contrast, the observer gain L has to be adapted to the new output matrix $C_{n,1}$ such that the prescribed eigenvalues of $A_n - LC_n$ are also the eigenvalues of $A_n - L_1 C_{n,1}$. A relation between L and L_1 can be derived by help of

the observer dynamics that, by aid of $e_n(t) = x_n(t) - \hat{x}_n(t)$, (3.42)–(3.43), and (3.50), is given by

$$\dot{e}_n(t) = (A_n - L_1 C_{n,1}) e_n(t) - L_1(\hat{y}_o(t) - y_{n,1}(t)). \tag{3.52}$$

The transformed error $\tilde{e}_n(t) := (A_n - \mu I)^{-1} e_n(t)$ has thus the dynamics

$$\dot{\tilde{e}}_n(t) = (A_n - \mu I)^{-1}(A_n - L_1 C_{n,1}) e_n(t) - (A_n - \mu I)^{-1} L_1(\hat{y}_o(t) - y_{n,1}(t)), \tag{3.53}$$

which by use of $(A_n - \mu I)^{-1} A_n = A_n (A_n - \mu I)^{-1}$ (see [89, Problem III 6.2]) and (3.44)–(3.45) can be written as

$$\dot{\tilde{e}}_n(t) = (A_n - (A_n - \mu I)^{-1} L_1 C_n) \tilde{e}_n(t) - (A_n - \mu I)^{-1} L_1 y_{r,1}(t). \tag{3.54}$$

Defining

$$L_1 := (A_n - \mu I) L \tag{3.55}$$

as the new observer gain therein yields

$$\dot{\tilde{e}}_n(t) = (A_n - L C_n) \tilde{e}_n(t) - L y_{r,1}(t). \tag{3.56}$$

This shows that by use of the observer gain (3.55) the observer dynamics of the compensator (3.50)–(3.51) is the same as the observer dynamics of the compensator (2.168)–(2.169) that was designed for the control loop without output observer.

In the next subsection the spillover is analyzed, which is resulting from the excitation of the observation error dynamics (3.56) by the output $y_{r,1}$ of the residual dynamics Σ_r^1. This excitation signal apparently contains the contribution $y_{r,1}(t)$ of the residual dynamics to the reconstructed output $\hat{y}_o$ and thus is responsible for the observation spillover.

3.2.2 Analysis of the spillover reduction

In Section 2.4 the spectrum perturbation due to spillover was estimated for the classical compensator design scheme. Now, the spectrum perturbation for the control loop with the additional output observer is considered and the improvement determined in a quantitative way. The behavior of the closed-loop system consisting of the extended system and the observer-based compensator can be described in terms of the redefined composed state

$$x_{cl}(t) := \begin{bmatrix} \tilde{e}_n(t) \\ x_n(t) \\ x_r(t) \end{bmatrix}. \tag{3.57}$$

When the state feedback

$$u(t) = -K\hat{x}_n(t) = -Kx_n(t) + K(A_n - \mu I)\tilde{e}_n(t) \tag{3.58}$$

is applied, which results from $\hat{x}_n(t) = x_n(t) - e_n(t)$ and $e_n(t) = (A_n - \mu I)\tilde{e}_n(t)$, combining the dynamics of $\tilde{e}_n$ (see (3.56)), Σ_n^1 (see (3.42)–(3.44)), and Σ_r^1 (see (3.46)–(3.49)) yields

$$\Sigma_{cl}^1: \quad \dot{x}_{cl}(t) = \mathcal{A}_{cl,1}x_{cl}(t) + \mathcal{L}e_o(t), \qquad t > 0,\ x_{cl}(0) = x_{cl,0} \in X_{cl} \tag{3.59}$$

$$\dot{e}_o(t) = \mu e_o(t), \qquad t > 0,\ e_o(0) = e_{o,0} \in \mathbb{C}^m \tag{3.60}$$

with $X_{cl} = \mathbb{C}^n \oplus \mathbb{C}^n \oplus X_r$. Therein, the operators $\mathcal{A}_{cl,1}$ and $\mathcal{L}$ are given by

$$\mathcal{A}_{cl,1} = \begin{bmatrix} A_n - LC_n & 0 & -L\mathcal{C}_r(\mathcal{A}_r - \mu I)^{-1} \\ B_nK(A_n - \mu I) & A_n - B_nK & 0 \\ \mathcal{B}_rK(A_n - \mu I) & -\mathcal{B}_rK & \mathcal{A}_r \end{bmatrix} \tag{3.61}$$

$$\mathcal{L} = \begin{bmatrix} L \\ 0 \\ 0 \end{bmatrix}, \tag{3.62}$$

where $D(\mathcal{A}_{cl,1}) = \mathbb{C}^n \oplus \mathbb{C}^n \oplus D(\mathcal{A}_r)$ and $D(\mathcal{L}) = \mathbb{C}^m$. A comparison of $\mathcal{A}_{cl,1}$ and $\mathcal{A}_{cl,0}$ in (2.225) shows that $\mathcal{A}_{cl,1}$ can be decomposed into

$$\mathcal{A}_{cl,1} = \mathcal{A}_{cl,0} + \Delta_1 \tag{3.63}$$

with

$$\Delta_1 = \begin{bmatrix} 0 & 0 & -L\mathcal{C}_r(\mathcal{A}_r - \mu I)^{-1} \\ B_nK(A_n - \mu I) & 0 & 0 \\ \mathcal{B}_rK(A_n - \mu I) & 0 & 0 \end{bmatrix}. \tag{3.64}$$

Thus, the system operators $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$ (see (2.224)) and $\mathcal{A}_{cl,1} = \mathcal{A}_{cl,0} + \Delta_1$ of the control loop without and with output observer, respectively, distinguish themselves only by the perturbation operator. If $\Delta_1 = 0$, the eigenvalues of the system operator $\mathcal{A}_{cl,1}$ are the desired eigenvalues of $A_n - LC_n$, $A_n - B_nK$, and $\mathcal{A}_r$ due to the block diagonal structure of $\mathcal{A}_{cl,0}$. For $\Delta_1 \neq 0$ all these eigenvalues are perturbed which is again the consequence of spillover. Note, that the error e_o does not influence the spectrum of $\mathcal{A}_{cl,0}$ because of its homogeneous dynamics. Some insight concerning a quantitative measure of this perturbation can be gained from the observation that Δ_1 relates to Δ (see (2.225)) by

$$\Delta_1 = \Delta\Lambda \tag{3.65}$$

with

$$\Lambda = \begin{bmatrix} A_n - \mu I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & (\mathcal{A}_r - \mu I)^{-1} \end{bmatrix}. \tag{3.66}$$

It has been shown in Section 2.4 that the spectrum perturbation depends on the norm of the disturbing operator (see (2.226)). The following estimate for $\|\Delta_1\|$ is therefore useful for the analysis of the spectrum perturbation.

Theorem 3.2-3

Suppose that L_1 and L are related according to (3.55), and assume $\mu \in S$ with S according to (3.13). Then, the perturbation operators Δ and Δ_1, that correspond to the closed-loop systems Σ_{cl} and Σ_{cl}^1 without and with output observer, respectively, (see (2.174) and (3.59)–(3.60)) satisfy

$$\frac{\|\Delta_1\|}{\|\Delta\|} \leq \eta := \sqrt{\frac{\eta_1}{\eta_2}} \tag{3.67}$$

with

$$\eta_1 = \|A_n - \mu I\|, \qquad \eta_2 = \|(\mathcal{A}_r - \mu I)^{-1}\|^{-1}. \tag{3.68}$$

For the proof see Appendix A.9.

Remark 3.2-4 In case that A_n and $\mathcal{A}_r$ have orthogonal eigenvectors the constants η_1 and η_2 have a simple interpretation. In this case A_n and $\mathcal{A}_r$ can be shown to be normal which leads to

$$\eta_1 = \max_{\lambda \in \sigma(A_n)} |\lambda - \mu|, \qquad \eta_2 = \inf_{\lambda \in \sigma(\mathcal{A}_r)} |\lambda - \mu| \tag{3.69}$$

(see [89, Sec. I 6.6 and Sec. V 3.8]). Thus, η_1 describes the radius of the smallest disk with the origin as its center that contains the whole spectrum of the approximation that is shifted by the output observer eigenvalue μ, and η_2 is the largest radius of a disc, centered at the origin, that does not contain any of the shifted spectral points of the residual dynamics (see Figure 17). So, the ratio η_1/η_2 in (3.67) characterizes the "spectral distance" between $\sigma(A_n - \mu I)$ and $\sigma(\mathcal{A}_r - \mu I)$. Of course, the estimate (3.67) is useful only, if $\eta < 1$ holds so that a reduction of the perturbation operator norm is assured. Fortunately, this condition is satisfied in most applications since

the approximation contains typically the eigenvalues with the smallest absolute values while the eigenvalues of the residual dynamics have larger absolute values (see Subsection 2.3.1). Furthermore, it is possible to lower η by means of a residual mode filter (see [13]). By such a filter some of the modes of Σ_r with largest eigenvalue absolute values can be compensated so that these modes do not contribute to the spillover effect. In consequence, the corresponding eigenvalues of $\mathcal{A}_r$ need not be taken into account for computing the infimum in (3.69), which leads to a lower value of η. In this way the effectiveness of the approach can be increased essentially, in particular when several output observers are used (see the next subsection). ◀

A more precise characterization of the spillover reduction can be obtained by estimating the spectrum perturbation

$$d_1 := \sup_{\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{cl,0}+\Delta_1)} \inf_{\lambda_{cl} \in \sigma(\mathcal{A}_{cl,0})} |\tilde{\lambda}_{cl} - \lambda_{cl}| \tag{3.70}$$

of the closed-loop system. Such an estimate can be obtained by help of (2.226) wherein only the perturbation operator Δ has to be replaced by Δ_1, which is possible because Δ_1 is a bounded operator as Δ is. This yields the relation $d_1 \leq \tilde{d}_1$ with

$$\tilde{d}_1 := \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta_1\| = \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\|\frac{\|\Delta_1\|}{\|\Delta\|} = \tilde{d}\frac{\|\Delta_1\|}{\|\Delta\|}, \tag{3.71}$$

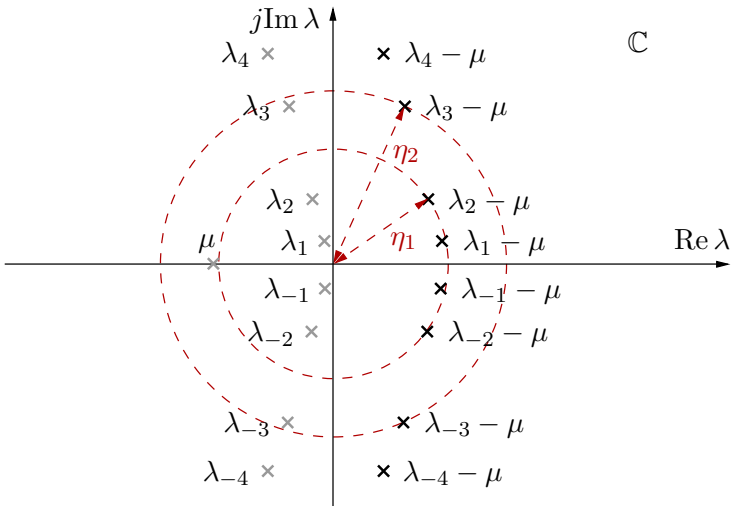

Figure 17 – Illustration of the spectral distance η_1/η_2 with $\eta_1 = \max_{\lambda\in\sigma(A_n)} |\lambda-\mu|$ and $\eta_2 = \inf_{\lambda\in\sigma(\mathcal{A}_r)} |\lambda-\mu|$ between the spectra $\sigma(A_n - \mu I) = \{\lambda_{\pm i} - \mu, i = 1, 2\}$ and $\sigma(\mathcal{A}_r - \mu I) = \{\lambda_{\pm i} - \mu, i \geq 3\}$.

in which $\mathcal{T}_{cl}$ is the normalizing transformation in the sense of Assumption 2.4-7. These considerations yield the following bound for d_1, that makes use of the Theorems 2.4-10 and 3.2-3.

Corollary 3.2-5

Consider the closed-loop system consisting of the plant Σ, the output observer Σ_o and the observer-based compensator Σ_c. Suppose that the following conditions are met:

1. *The Assumptions 2.1-9 and 2.4-7 hold,*
2. *the observer gains L_1 and L satisfy (3.55), and*
3. *it holds $\mu \in S$ (see (3.13)).*

Then, the spectrum perturbation d_1 in (3.70) of the closed-loop system has the upper bound

$$d_1 \leq \tilde{d}_1 \quad \textit{with} \quad \tilde{d}_1 = \tilde{d}\,\eta, \tag{3.72}$$

where η is given by (3.67)–(3.68) and $\tilde{d} = \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\|$ (see (2.226)).

Remember that $\tilde{d}$ is an upper bound for the spectrum perturbation d that corresponds to the control loop without any output observer (see Section 2.4). In contrast, $\tilde{d}_1$ is an upper bound for the spectrum perturbation d_1 of the control system that utilizes an output observer. According to (3.72) this upper bound is reduced compared to $\tilde{d}$ if $\eta < 1$. Thus, when the spectrum perturbation is considered as a measure of the spillover, this makes apparent that the spectrum perturbation is suppressed whenever η is small, which in fact is satisfied in many applications as explained above (see Remark 3.2-4). Hence, (3.72) characterizes the benefit of inserting the output observer into the control loop.

3.2.3 Improved spillover reduction by cascaded output observers

The full strength of the approach presented so far lies in the fact that it can be applied in a repeated manner for successively reducing the spillover. To do so, an additional output observer is added to the extended system (3.37)–(3.39) as shown in Figure 18 (see also Figure 14b). Therein, Σ_o^1 equals the output observer Σ_o that was used already in the subsection before, and Σ_o^2 is an additional output observer. This one is designed

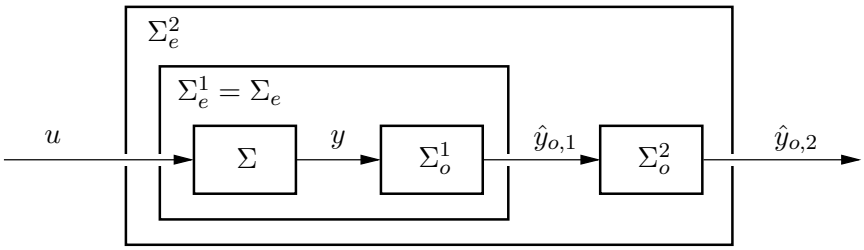

Figure 18 – Structure of the system extension with two output observers $\Sigma_o^1 := \Sigma_o$ and Σ_o^2.

in the same way as Σ_o^1 with the only difference that $\Sigma_e^1 := \Sigma_e$ is taken as the basis for its design instead of Σ. The observer-based compensator uses the output $\hat{y}_{o,2}$ of the doubly extended system and is designed for an approximation of it. It suggests itself to add output observers repeatedly since in this way the spectrum perturbation of the closed-loop system operator can be successively reduced (see Figure 14). In what follows, it is assumed that $q > 1$ output observers are used. In order to describe the cascaded control system structure adequately some notation has to be introduced.

Definition 3.2-6

Let $\Sigma_e^i := (\Sigma_e^{i-1}, \Sigma_o^i)$, $i = 1, 2, \ldots, q$, denote the composition of a system Σ_e^{i-1} and the associated output observer Σ_o^i, wherein $\Sigma_e^0 := \Sigma$ and $\Sigma_o^1 := \Sigma_o$. The fictitious output of Σ_e^i is denoted $y_{o,i}$ for that the output observer Σ_o^i provides the asymptotic estimation $\hat{y}_{o,i}$. Furthermore, Σ_n^i stands for the n-dimensional modal approximation of Σ_e^i with output $y_{n,i}$, and Σ_r^i is the corresponding residual dynamics with output $y_{r,i}$ that satisfies

$$y_{n,i}(t) + y_{r,i}(t) = \hat{y}_{o,i}(t), \qquad i = 1, 2, \ldots, q. \tag{3.73}$$

Finally, Σ_c^i denotes the observer-based compensator designed on the basis of Σ_n^i with controller gain K and observer gain L_i. ◀

As before, the output observers have the form

$$\Sigma_o^i: \quad \dot{\hat{y}}_{o,i}(t) = \mu \hat{y}_{o,i}(t) + B_{o,i} u(t) + \hat{y}_{o,i-1}(t), \qquad t > 0, \quad \hat{y}_{o,i}(0) \in \mathbb{C}^m, \tag{3.74}$$

for $i = 1, 2, \ldots, q$, where $\hat{y}_{o,0}(t) := y(t)$ (compare to (3.41)). They are intended to reconstruct the fictitious output $y_{o,i}(t)$ asymptotically. The same reasoning as in Section 3.1 reveals that the fictitious outputs, that can be reconstructed by the output

observers, have the form $y_{o,i}(t) = \mathcal{C}_i x(t)$ with

$$\mathcal{C}_i = \mathcal{C}(\mathcal{A} - \mu I)^{-i}. \tag{3.75}$$

In Section 3.1 it was found that the fictitious output $y_{o,1}$ contains less contributions of the residual modes than the system output y does, which leads to reduced spillover when $y_{o,1}$ is used for the compensator. This effect can be understood in a way that the factor $(\mathcal{A} - \mu I)^{-1}$ in $\mathcal{C}_1 = \mathcal{C}(\mathcal{A} - \mu I)^{-1}$ deals as a filter which suppresses the residual modes. For that reason it is clear that using the reconstructed output $y_{o,q} = \mathcal{C}_q x = \mathcal{C}(\mathcal{A} - \mu I)^{-q} x$ (see (3.75)) for the compensator allows to successively reduce the spillover by increasing the number q of output observers.

For discussing the dynamics of the observation errors

$$e_{o,i}(t) \coloneqq y_{o,i}(t) - \hat{y}_{o,i}(t) = \mathcal{C}_i x(t) - \hat{y}_{o,i}(t), \qquad i = 1, 2, \ldots, q \tag{3.76}$$

note, that in view of (3.74) all output observers have the same eigenvalue $\mu \in S = (-\infty, 0) \cap \rho(\mathcal{A})$, for simplicity. While Σ_o^1 was designed in Section 3.1 such that the error $e_{o,1}$ satisfied the homogeneous dynamics $\dot{e}_{o,1}(t) = \mu e_{o,1}(t)$ (see Theorem 3.1-3), the further output observers Σ_o^i, $i > 1$, have different error dynamics because the error of an output observer Σ_o^i is fed into the following output observer Σ_o^{i+1}. By the same considerations as in Section 3.1 it can be shown that the matrix $B_{o,i}$ in (3.74), that corresponds to the output operator (3.75), is given by

$$B_{o,i} = \mathcal{C}_i \mathcal{B}, \tag{3.77}$$

and the dynamics of the observer errors read

$$\dot{e}_{o,i}(t) = \mu e_{o,i}(t) + e_{o,i-1}(t), \qquad i > 1. \tag{3.78}$$

Thus, the vector

$$\bar{e}_o \coloneqq \begin{bmatrix} e_{o,1} \\ e_{o,2} \\ \vdots \\ e_{o,q} \end{bmatrix} \tag{3.79}$$

of the errors of all output observers satisfies

$$\dot{\bar{e}}_o(t) = M \bar{e}_o(t) \tag{3.80}$$

with

$$M = \begin{bmatrix} \mu & & & 0 \\ 1 & \mu & & \\ & \ddots & \ddots & \\ 0 & & 1 & \mu \end{bmatrix}. \tag{3.81}$$

Since M is a Hurwitz matrix due to $\mu \in S$, the error $\bar{e}_o$ decays exponentially toward zero so that the fictitious outputs $y_{o,i}$, $i = 1, 2, \ldots, q$, are indeed reconstructed asymptotically as intended.

For the compensator design the approximation

$$\Sigma_n^q: \quad \dot{x}_n(t) = A_n x_n(t) + B_n u(t), \qquad t > 0,\ x_n(0) = \mathcal{F}^{-1}\mathcal{P}x_0 \in \mathbb{C}^n \tag{3.82}$$

$$y_{n,q}(t) = C_{n,q} x_n(t), \qquad t \geq 0 \tag{3.83}$$

with

$$C_{n,q} = C_n (A_n - \mu I)^{-q} \tag{3.84}$$

is considered which is the modal approximation of Σ_e^q that contains the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathcal{A}$. The residual dynamics

$$\Sigma_r^q: \quad \dot{x}_r(t) = \mathcal{A}_r x_r(t) + \mathcal{B}_r u(t), \qquad t > 0,\ x_r(0) = (I - \mathcal{P})x_0 \in X_r \tag{3.85}$$

$$\dot{\bar{e}}_o(t) = M\bar{e}_o(t), \qquad t > 0,\ \bar{e}_o(0) = \bar{e}_{o,0} \in \mathbb{C}^{qm} \tag{3.86}$$

$$y_{r,q}(t) = \mathcal{C}_{r,q} x_r(t) - \begin{bmatrix} 0 & \cdots & 0 & I \end{bmatrix} \bar{e}_o(t), \quad t \geq 0 \tag{3.87}$$

with

$$\mathcal{C}_{r,q} = \mathcal{C}_r (\mathcal{A}_r - \mu I)^{-q} \tag{3.88}$$

describes the output error

$$y_{r,q}(t) = \hat{y}_{o,q}(t) - y_{n,q}(t) \tag{3.89}$$

of the approximation Σ_n^q, which can be verified in the same way as Lemma 3.2-1. Next, the observer-based compensator on the basis of Σ_n^q is designed that reads

$$\Sigma_c^q: \quad \dot{\hat{x}}_n(t) = (A_n - L_q C_{n,q})\hat{x}_n(t) + B_n u(t) + L_q \hat{y}_{o,q}(t), \quad t > 0,\ \hat{x}_n(0) \in \mathbb{C}^n \tag{3.90}$$

$$u(t) = -K\hat{x}_n(t), \qquad t \geq 0. \tag{3.91}$$

The corresponding error $\tilde{e}_n(t) \coloneqq (A_n - \mu I)^{-q} e_n(t)$ has the dynamics

$$\dot{\tilde{e}}_n(t) = (A_n - (A_n - \mu I)^{-q} L_q C_n)\tilde{e}_n(t) - (A_n - \mu I)^{-q} L_q y_{r,q}(t) \tag{3.92}$$

which is the analog of (3.54). Apparently, the observer dynamics is the same as the dynamics of the observer in the control system without output observer if

$$L_q \coloneqq (A_n - \mu I)^q L \tag{3.93}$$

is used as the new observer gain because the dynamic matrix in (3.92) simplifies then to $A_n - (A_n - \mu I)^{-q} L_q C_n = A_n - LC_n$. By combing the dynamics of $\tilde{e}_n$, Σ_n^q, and Σ_r^q

(see (3.92), (3.82)–(3.84), and (3.85)–(3.88)) the closed-loop dynamics in terms of the state

$$x_{cl}(t) = \begin{bmatrix} \tilde{e}_n(t) \\ x_n(t) \\ x_r(t) \end{bmatrix} \tag{3.94}$$

reads

$$\Sigma_{cl}^q: \quad \dot{x}_{cl}(t) = (\mathcal{A}_{cl,0} + \Delta_q)x_{cl}(t) + \mathcal{L}\bar{e}_o(t), \qquad t > 0,\ x_{cl}(0) = x_{cl,0} \in X_{cl} \tag{3.95}$$

$$\dot{\bar{e}}_o(t) = M\bar{e}_o(t), \qquad t > 0,\ \bar{e}_o(0) = \bar{e}_{o,0} \in \mathbb{C}^{qm} \tag{3.96}$$

with $\mathcal{A}_{cl,0}$ according to (2.225), M according to (3.81), and

$$\mathcal{L} = \begin{bmatrix} 0 & \cdots & 0 & L \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \tag{3.97}$$

$$\Delta_q = \begin{bmatrix} 0 & 0 & -L\mathcal{C}_r(\mathcal{A}_r - \mu I)^{-q} \\ B_n K(A_n - \mu I)^q & 0 & 0 \\ \mathcal{B}_r K(\mathcal{A}_r - \mu I)^q & 0 & 0 \end{bmatrix} = \Delta\,\Lambda^q, \tag{3.98}$$

wherein the right equation for Δ_q follows from the comparison with (2.225) and (3.66). As before, the part $\mathcal{A}_{cl,0}$ of the system operator in (3.95) has the desired eigenvalues of $A_n - LC_n$, $A_n - B_nK$, and $\mathcal{A}_r$ that are perturbed by the operator Δ_q, while the error $\bar{e}_o$ does not influence the spectrum of $\mathcal{A}_{cl,0}$ because of its homogeneous dynamics. As argued in the previous subsections, the spectrum perturbation due to spillover depends on the norm of Δ_q which is analyzed next. From (3.93) it follows $L_q = (A_n - \mu I)L_{q-1}$, and (3.98) gives $\Delta_q = \Delta_{q-1}\Lambda$. Therefore, Theorem 3.2-3 can be applied when L, L_1, Δ, and Δ_1 are replaced by L_{q-1}, L_q, Δ_{q-1}, and Δ_q, respectively, within the theorem. This yields the estimate $\|\Delta_q\|/\|\Delta_{q-1}\| \leq \eta$ with η according to (3.67)–(3.68), and using this result recursively leads to

$$\frac{\|\Delta_q\|}{\|\Delta\|} \leq \eta^q. \tag{3.99}$$

In case that η is smaller than 1 this shows that the norm $\|\Delta_q\|$ of the perturbation operator decreases rapidly by increasing the number q of output observers. One can therefore expect that also the closed-loop spectrum perturbation

$$d_q := \sup_{\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{cl,0} + \Delta_q)} \inf_{\lambda_{cl} \in \sigma(\mathcal{A}_{cl,0})} |\tilde{\lambda}_{cl} - \lambda_{cl}| \tag{3.100}$$

is successively reduced by repeatedly adding output observers. In order to derive an upper bound $\tilde{d}_q$ for d_q (2.226) is applied with Δ replaced by Δ_q. This is possible

because the system operator of Σ_{cl}^q has the same unperturbed part $\mathcal{A}_{cl,0}$ as the system operator of the closed-loop system Σ_{cl} without any output observer (compare (3.95) and (2.174) using (2.224)), and Δ_q is a bounded operator as Δ is. Doing so yields $d_q \leq \tilde{d}_q$ with

$$\tilde{d}_q := \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta_q\| = \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\|\frac{\|\Delta_q\|}{\|\Delta\|} = \tilde{d}\frac{\|\Delta_q\|}{\|\Delta\|}, \tag{3.101}$$

in which $\mathcal{T}_{cl}$ is the normalizing transformation in the sense of Assumption 2.4-7. Using this result the following upper bound for d_q is obtained from Theorem 2.4-10 and Theorem 3.2-3.

Corollary 3.2-7

Consider the closed-loop system consisting of the plant Σ, the q output observers Σ_o^i, $i = 1, \ldots, q$, and the observer-based compensator Σ_c^q. Suppose that the following conditions are met:

1. *The Assumptions 2.1-9 and 2.4-7 hold,*
2. *the observer gains L_q and L satisfy (3.93), and*
3. *all output observers have the same eigenvalue $\mu \in S$ (see (3.13)).*

Then, the spectrum perturbation d_q of the closed-loop system has the upper bound $d_q \leq \tilde{d}_q$ with

$$\tilde{d}_q = \tilde{d}\,\eta^q, \tag{3.102}$$

wherein η is given by (3.67)–(3.68) and $\tilde{d} = \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\|$ (see (2.226)).

Relation (3.102) makes apparent that the bound $\tilde{d}_q$ for the spectrum perturbation d_q decreases *exponentially* toward zero with respect to q provided that $\eta < 1$ holds. For that reason a certain degree of reduction of the spectrum perturbation can be achieved in many cases with a smaller compensator order compared to the conventional early-lumping design. If an admissible perturbation d_q is given and $\tilde{d}$ is known, then (3.102) can be solved for the number

$$q \geq \log_\eta \frac{d_q}{\tilde{d}} = \frac{\ln(d_q/\tilde{d})}{\ln(\eta)} \tag{3.103}$$

of output observers that are needed in the worst case to guarantee the specified spectrum perturbation d_q. Since each output observer has the order m by assumption, the

required compensator order is $n_c = n + mq$ which hence can be determined *a priori*. When, for comparison, a classically designed observer-based compensator is used, an estimate is not available that relates the order n_c to the spectrum perturbation or vice versa. Furthermore, the applications reveal that a significant decrease of the spectrum perturbation requires often a comparatively high compensator order so that the required efforts for design and implementation may be undesired high.

Remark 3.2-8 A smaller compensator order can be achieved when the assumption that the order of the output observers satisfies $n_o = m$ is removed. The output observers have then the more general form (3.2) with the extra parameter $L_o \in \mathbb{C}^{n_o \times m}$. The degrees of freedom in L_o can be used for additionally reducing the spillover because the impact of $m - 1$ modal states of the residual dynamics Σ_r can be eliminated by choosing L_o appropriately. ◀

Remark 3.2-9 For the evaluation of (3.102) and (3.103) $\tilde{d} = \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\|$ (see (2.226)) is needed. In fact, since Δ has finite rank, it is possible to compute $\|\Delta\|$ exactly by means of standard software tools, and for $\|\mathcal{T}_{cl}\|$ and $\|\mathcal{T}_{cl}^{-1}\|$ some easy to evaluate upper bounds can be derived when it is taken into account that $\mathcal{T}_{cl}$ and $\mathcal{T}_{cl}^{-1}$ have a special form due to the triangular structure of $\mathcal{A}_{cl,0}$. However, the constant $\tilde{d}$ leads often to very conservative estimates (3.102). This relation is therefore primarily of qualitative value. Most applications reveal that the actual spillover reduction is much better than guaranteed by this estimate. This turns out also in the following example. ◀

Example 3.2-10 (Euler-Bernoulli beam with Kelvin-Voigt damping, continued)

The presented approach is applied to the control of the Euler-Bernoulli beam with Kelvin-Voigt damping for that a state space model (2.3)–(2.4) with

$$\mathcal{A}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} \mathcal{A}_0 h_2 \\ -\mathcal{A}_0(h_1 + 2\delta\mathcal{A}_0 h_2) \end{bmatrix}, \qquad \forall \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \in D(\mathcal{A}) \tag{3.104}$$

$$\mathcal{B}v = \begin{bmatrix} 0 \\ b \end{bmatrix}, \qquad \forall\, v \in \mathbb{C} \tag{3.105}$$

$$\mathcal{C}h = \left\langle h, \begin{bmatrix} \mathcal{A}_0^{-1} c \\ 0 \end{bmatrix} \right\rangle_X, \qquad \forall\, h \in X \tag{3.106}$$

was derived in Example 2.1-15, where $\mathcal{A}_0$ is given in (2.53). The parameters that determine the input and output distribution functions $b(z) = \frac{1}{\beta_2-\beta_1} \cdot \mathbf{1}_{[\beta_1,\beta_2]}$ and $c(z) = \frac{1}{\gamma_2-\gamma_1} \cdot \mathbf{1}_{[\gamma_1,\gamma_2]}$, respectively, are $\beta_1 = 0.199$, $\beta_2 = 0.201$, $\gamma_1 = 0.749$, and $\gamma_2 = 0.701$ in this example, and the length and the damping constant are $\ell = 1$ and $\delta = 10^{-3}$, respectively. It has been found in Example 3.1-4 that the SDGA for the closed-loop system operator is satisfied which is preliminary for a compensator design by eigenvalue assignment to make sense. Since the pairs of eigenvalues $\lambda_{\pm 1} = -0.097 \pm j9.9$, $\lambda_{\pm 2} = -1.6 \pm j39.4$ and $\lambda_{\pm 3} = -7.9 \pm j88.5$ are badly-damped with damping constants $D_{\pm 1} = 0.0099$, $D_{\pm 2} = 0.040$, and $D_{\pm 3} = 0.089$, respectively, it is desired to shift these eigenvalues by means of an observer-based compensator that is designed on the basis of a modal approximation Σ_n of the order $n = 6$ that contains these modes. In order to determine A_n, B_n, and C_n of the approximation by aid of (2.108)–(2.110), the biorthonormal sequence $\{\psi_i, i \geq 1\}$ associated to the eigenvectors $\{\phi_i, i \geq 1\}$ according to (2.71)–(2.72) is needed. It is straightforward to verify that this sequence is given by

$$\psi_i(z) = \sin(i\pi z) \frac{-2(i\pi)^4}{\overline{\lambda_i} - \overline{\lambda_{-i}}} \begin{bmatrix} \frac{\overline{\lambda_{-i}}}{(i\pi)^4} \\ \frac{1}{(i\pi)^2} \end{bmatrix}, \qquad i \in \mathbb{N} \tag{3.107}$$

$$\psi_{-i}(z) = \sin(i\pi z) \frac{2(i\pi)^4}{\overline{\lambda_i} - \overline{\lambda_{-i}}} \begin{bmatrix} \frac{\overline{\lambda_i}}{(i\pi)^2} \\ 1 \end{bmatrix}, \qquad i \in \mathbb{N} \tag{3.108}$$

so that the approximation can be computed by aid of (2.108)–(2.110). Next, the observer-based compensator Σ_c (see (2.168)–(2.169)), that uses the available measurement y, is designed by eigenvalue assignment. It is desired to achieve a stability margin $\beta \geq 5$ and a damping $D \geq \frac{1}{2}\sqrt{2}$. The eigenvalues $\lambda_{c,i}$ that are assigned to $A_n - B_n K$ and the eigenvalues $\lambda_{o,i}$ assigned to $A_n - LC_n$ are listed in Table 4. Unfortunately, the resulting closed-loop system consisting of the plant Σ and the compensator Σ_c has an unstable complex conjugate pair of eigenvalues $\tilde{\lambda}_{cl,\pm 1} = 0.15 \pm j5.53$ due to the spillover. For its suppression q output observers $\Sigma_o^1, \ldots, \Sigma_o^q$ according to (3.41) are added, where, in doing so, their eigenvalues are assigned to $\mu = -12$. The resulting distribution of

Table 4 – Eigenvalues $\lambda_{c,i}$ assigned to $A_n - B_n K$ and $\lambda_{o,i}$ assigned to $A_n - LC_n$.

i	$1, 2$	$3, 4$	$5, 6$
$\lambda_{c,i}$	$-8 \pm j2$	$-23 \pm j14$	$-39 \pm j28$
$\lambda_{o,i}$	$-12 \pm j2$	$-27 \pm j14$	$-43 \pm j28$

the seven most dominant pairs of closed-loop eigenvalues $\tilde{\lambda}_{cl,\pm 1}, \tilde{\lambda}_{cl,\pm 2}, \ldots, \tilde{\lambda}_{cl,\pm 7}$ are displayed for $q = 4$ in Figure 19, together with the desired eigenvalue locations and the boundary of the specified eigenvalue region. The computation of the closed-loop eigenvalues has been carried out on the basis of a 60^{th}-order modal approximation of the closed-loop system. For comparison, a residual mode filter is applied instead of the cascade of output observers. The filter order is $n_{\text{rmf}} = 4$ so that the order of the dynamic system consisting of the observer-based compensator and the residual mode filter coincides with the order of the compensator with the $q = 4$ output observers, since each of which has the order $n_o = 1$. The resulting eigenvalue distribution is also shown in Figure 19. Apparently, the residual mode filter does not yield satisfying closed-loop dynamics because there are several pairs of eigenvalues outside the specified region. In fact, the eigenvalues are not placed within that region unless n_{rmf} is increased up to $n_{\text{rmf}} \geq 23$. The decrease of the spectrum perturbation d_q by increasing q and n_{rmf} is illustrated in Figure 20, where the mentioned modal approximation of the beam with order $n_{\text{high}} = 60$ has been used for the computations. The figure shows that the effect of the residual mode filter is moderate even for a comparatively high order n_{rmf} while the presented approach allows to lower d_q effectively. According to Theorem 3.2-3 one obtains $\eta = \|\Delta_1\|/\|\Delta\| = 0.53 < 1$, where it is taken into account that the eigenvalues $\lambda_{\pm 4}$ and $\lambda_{\pm 5}$ are not modal observable and not modal controllable, respectively, so that these can be omitted in the evaluation of (3.68) (see Remark 3.2-4). It has been found in Example 2.1-15 that Assumption 2.1-9 is satisfied and Assumption 2.4-7 holds in view of Proposition 2.4-8 because $\mathcal{A}$ was shown in Example 2.1-15 to be a Riesz-spectral operator and the eigenvalues of $A_n - B_n K$ and $A_n - LC_n$ have been assigned appropriately (see Table 4 and (2.68)). Thus, Corollary 3.2-7 can be applied with the observer gain $L_q = (A_n - \mu I)^q L$ so that relation (3.103) can be used. It reveals that the number $q = 16$ of output observers assures closed-loop stability of the transformed system which is equivalent to the stabilization of the original system. However, this estimate is quite conservative. In fact, for achieving the specified minimum stability margin and minimum damping $q = 4$ output observers are sufficient. ◀

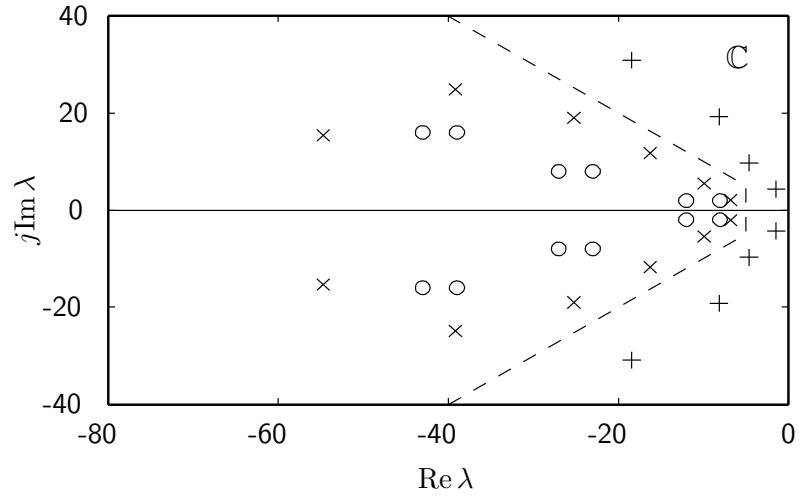

Figure 19 – Desired eigenvalue locations ('o'), eigenvalues resulting from the use of $q = 4$ output observers ('x') and from the use of a residual mode filter of the order $n_{\mathrm{rmf}} = 4$ ('+'). The dashed lines define the boundary of the region with the required stability margin $\beta \geq 5$ and damping $D \geq \frac{1}{2}\sqrt{2}$.

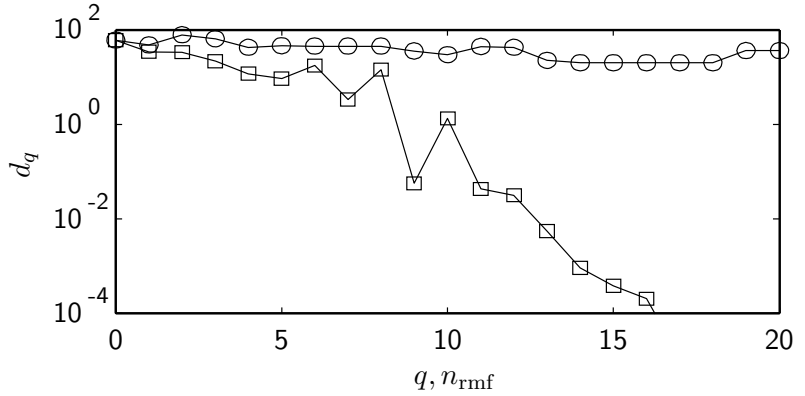

Figure 20 – Spectrum perturbation d_q of the closed-loop system with q output observers of order $n_o = 1$ ('□') and with residual mode filter of the order n_{rmf} ('○'). The spectrum perturbation has been computed on the basis of a modal approximation of the beam with order $n_{\mathrm{high}} = 60$.

3.3 Observation spillover reduction versus control spillover reduction

It has been explained in Subsection 2.3.2 that the overall spillover is caused by two different effects: first, the observation spillover, by which the impact of the residual dynamics on the observer dynamics

$$\dot{e}_n(t) = (A_n - LC_n)e_n(t) - L\mathcal{C}_r x_r(t) \tag{3.109}$$

(see Section 2.3.2) is meant, where $e_n = x_n - \hat{x}_n$ is the observation error. The second spillover effect comes from the excitation of the residual dynamics by the observation error, which is called control spillover and is described by

$$\dot{x}_r(t) = \mathcal{A}_r x_r(t) - \mathcal{B}_r K x_n(t) + \mathcal{B}_r K e_n(t) \tag{3.110}$$

(see (2.171)). Since the term $L\mathcal{C}_r x_r(t) = Ly_r(t)$ in (3.109) is responsible for the observation spillover it is the basic idea of the compensator design approach in the previous Section 3.2 to utilize reconstructed outputs for the compensator. In consequence, the mentioned term is modified in an advantageous way. In fact, the observer dynamics (3.109) becomes

$$\dot{e}_n(t) = (A_n - L_q C_n(A_n - \mu I)^{-q})e_n(t) - L_q\mathcal{C}_r(\mathcal{A}_r - \mu I)^{-q}x_r(t) + L_q e_{o,q}(t) \tag{3.111}$$

for the control loop with q output observers, which follows from (3.82)–(3.84), and (3.87)–(3.90). The comparison with (3.109) shows that the excitation of e_n by the residual state x_r has changed due to the new output operator $\mathcal{C}_r(\mathcal{A}_r - \mu I)^{-q}$ in (3.111). Thus, the spillover reduction of the design approach is achieved by modifying the observation spillover. In contrast, the control spillover is unaffected since (3.110) remains valid for the control loop with output observers (see (3.85) and (3.91)). This raises the following questions:

1. Is it possible to suppress the control spillover alternatively or additionally to the observation spillover?
2. How is a modification of the control spillover related to the corresponding spectrum perturbation?

These issues are discussed in the following two subsections.

3.3.1 Modification of the control spillover

Here, the possibility to affect the control spillover instead of or in addition to the observation spillover is analyzed. For this purpose the residual dynamics

$$\Sigma_r^q: \quad \dot{x}_r(t) = \mathcal{A}_r x_r(t) + \mathcal{B}_r u(t), \qquad t > 0,\ x_r(0) = x_{r,0} \in X_r \tag{3.112}$$

$$\dot{\bar{e}}_o(t) = M\bar{e}_o(t), \qquad t > 0,\ \bar{e}_o(0) = \bar{e}_{o,0} \in \mathbb{C}^{qm} \tag{3.113}$$

$$y_{r,q}(t) = \mathcal{C}_r(\mathcal{A}_r - \mu I)^{-q} x_r(t) - \bar{e}_o(t), \qquad t \geq 0 \tag{3.114}$$

of the extended system (see (3.85)–(3.88)), that corresponds to the closed-loop system with q output observers, is rewritten in terms of the transformed state

$$\tilde{x}_r(t) \coloneqq (\mathcal{A}_r - \mu I)^{-k} x_r(t), \qquad k \in \{0, 1, \ldots, q\}, \tag{3.115}$$

in which $(\mathcal{A}_r - \mu I)^{-k}$ is defined on whole X_r because $\mu \in \rho(\mathcal{A}) \subset \rho(\mathcal{A}_r)$ was claimed before (see (3.13)). This leads to

$$\tilde{\Sigma}_r^q: \quad \dot{\tilde{x}}_r(t) = \mathcal{A}_r \tilde{x}_r(t) + (\mathcal{A}_r - \mu I)^{-k} \mathcal{B}_r u(t), \qquad t > 0,\ \tilde{x}_r(0) = \tilde{x}_{r,0} \in X_r \tag{3.116}$$

$$\dot{\bar{e}}_o(t) = M\bar{e}_o(t), \qquad t > 0,\ \bar{e}_o(0) = \bar{e}_{o,0} \in \mathbb{C}^{qm} \tag{3.117}$$

$$y_{r,q}(t) = \mathcal{C}_r(\mathcal{A}_r - \mu I)^{k-q} \tilde{x}_r(t) - \bar{e}_o(t), \qquad t \geq 0, \tag{3.118}$$

where it is used that an operator and its resolvent commutes[2] (see [89, Problem III 6.2]) so that

$$(\mathcal{A}_r - \mu I)^{-k} \mathcal{A}_r (\mathcal{A}_r - \mu I)^k = \mathcal{A}_r (\mathcal{A}_r - \mu I)^{-k} (\mathcal{A}_r - \mu I)^k = \mathcal{A}_r \tag{3.119}$$

holds. Apparently, the state transformation has the effect that both the input and the output operator become changed while the system operator remains unaffected. Similarly, the approximation of the extended system with q output observers, which is described by

$$\Sigma_n^q: \quad \dot{x}_n(t) = A_n x_n(t) + B_n u(t), \qquad t > 0,\ x_n(0) = x_{n,0} \in \mathbb{C}^n \tag{3.120}$$

$$y_{n,q}(t) = C_n(A_n - \mu I)^{-q} x_n(t), \qquad t \geq 0 \tag{3.121}$$

(see (3.82)–(3.84)), is expressed by the transformed state

$$\tilde{x}_n(t) \coloneqq (A_n - \mu I)^{-k} x_n(t), \qquad k \in \{0, 1, \ldots, q\}, \tag{3.122}$$

[2] Two operators $\mathcal{M}_1 : D(\mathcal{M}_1) = H \mapsto H$, $\mathcal{M}_2 : D(\mathcal{M}_2) \subseteq H \mapsto H$ are said to *commute* if $\mathcal{M}_1\mathcal{M}_2 \subset \mathcal{M}_2\mathcal{M}_1$ holds, *i.e.*, $\mathcal{M}_1 h \in D(\mathcal{M}_2)$, $\forall\, h \in D(\mathcal{M}_2)$, and $\mathcal{M}_2\mathcal{M}_1 h = \mathcal{M}_1\mathcal{M}_2 h$, $\forall\, h \in D(\mathcal{M}_2)$.

where the existence of the inverse follows again from $\mu \in \rho(\mathcal{A}) \subset \rho(A_n)$, yielding

$$\tilde{\Sigma}_n^q: \quad \dot{\tilde{x}}_n(t) = A_n\tilde{x}_n(t) + (A_n - \mu I)^{-k}B_n u(t), \quad t > 0,\ \tilde{x}_n(0) = \tilde{x}_{n,0} \in \mathbb{C}^n \tag{3.123}$$

$$y_{n,q}(t) = C_n(A_n - \mu I)^{k-q}\tilde{x}_n(t), \qquad t \geq 0. \tag{3.124}$$

By insertion of $x_n(t) = (A_n - \mu I)^k \tilde{x}_n(t)$ and $x_r(t) = (\mathcal{A}_r - \mu I)^k \tilde{x}_r(t)$ into (3.110)–(3.111) one arrives at

$$\dot{\tilde{x}}_r(t) = \mathcal{A}_r\tilde{x}_r(t) - (\mathcal{A}_r - \mu I)^{-k}\mathcal{B}_r K(A_n - \mu I)^k\tilde{x}_n(t) + (\mathcal{A}_r - \mu I)^{-k}\mathcal{B}_r K e_n(t) \tag{3.125}$$

$$\dot{e}_n(t) = (A_n - L_q C_n(A_n - \mu I)^{-q})e_n(t) - L_q\mathcal{C}_r(\mathcal{A}_r - \mu I)^{k-q}\tilde{x}_r(t) + L_q e_{o,q}(t). \tag{3.126}$$

This shows that the operators $\mathcal{B}_r$ and $\mathcal{C}_r(\mathcal{A}_r - \mu I)^{-q}$ in (3.110)–(3.111), that are related to the control and observation spillover, respectively, are replaced by $(\mathcal{A}_r - \mu I)^{-k}\mathcal{B}_r$ and $\mathcal{C}_r(\mathcal{A}_r - \mu I)^{k-q}$ in (3.125)–(3.126). Thus, the state transformations have the consequence that both the control and the observation spillover are changed, at which the parameter k controls the individual change of both kinds. However, it will be shown in the next subsection that it is not possible to suppress both sorts of spillover at the same time. Instead, the parameter k determines if the control or the observation spillover is more dominant while the other is suppressed. For the particular choice $k = q$ the operator $\mathcal{C}_r(\mathcal{A}_r - \mu I)^{k-q}$ in (3.126) simplifies to $\mathcal{C}_r(\mathcal{A}_r - \mu I)^{q-q} = \mathcal{C}_r$ so that the observer dynamics is influenced by the same output of the residual dynamics as in the control loop without any output observers (see (3.109)). Thus, the compensator design approach in Section 3.2 yields in this case a modification of the control spillover the observation spillover is not affected. For $k = 0$, in contrast, one has the situation discussed in Subsection 3.2.3, where the utilization of the output observers suppressed solely the observation spillover but not the control spillover. Finally, by choosing k in between these two special cases the impact of the fictitious outputs is divided up between both kinds of spillover. However, the question remains whether for any choice of k the same extent of reduction of the spectrum perturbation is achieved. This is investigated next.

3.3.2 Reduction of the spectrum perturbation by modification of the control spillover

Since only state transformations have been used in the considerations before it can be expected that modifying the control spillover by introducing fictitious outputs and

choosing $k \in \{1, 2, \ldots, q\}$ has the same effect on the spectrum perturbation as it was shown in Subsection 3.2.3, where the observation spillover was modified. This is proven in the following by considering the closed-loop dynamics in terms of the state

$$\tilde{x}_{cl}(t) := \begin{bmatrix} \tilde{e}_n(t) \\ \tilde{x}_n(t) \\ \tilde{x}_r(t) \end{bmatrix} \tag{3.127}$$

that is composed of the transformed states $\tilde{e}_n(t) = (A_n - \mu I)^{-q} e_n(t)$ and $\tilde{x}_n$ and $\tilde{x}_r$ from (3.122) and (3.115), respectively. It is straightforward to show that the dynamics of the control loop with q output observers with regard to $\tilde{x}_{cl}$ is described by

$$\tilde{\Sigma}_{cl}^{q}: \quad \dot{\tilde{x}}_{cl}(t) = \tilde{\mathcal{A}}_{cl,q}\tilde{x}_{cl}(t) + \mathcal{L}\bar{e}_o(t), \qquad t > 0, \quad \tilde{x}_{cl}(0) = \tilde{x}_{cl,0} \in X_{cl} \tag{3.128}$$

$$\dot{\bar{e}}_o(t) = M\bar{e}_o(t), \qquad t > 0, \quad \bar{e}_o(0) = \bar{e}_{o,0} \in \mathbb{C}^{qm} \tag{3.129}$$

with $X_{cl} = \mathbb{C}^n \oplus \mathbb{C}^n \oplus X_r$, M according to (3.81), and

$$\tilde{\mathcal{A}}_{cl,q} = \begin{bmatrix} A_n - LC_n & 0 & -L\mathcal{C}_r(\mathcal{A}_r - \mu I)^{k-q} \\ (A_n - \mu I)^{-k} B_n K (A_n - \mu I)^q & A_n - (A_n - \mu I)^{-k} B_n K (A_n - \mu I)^k & 0 \\ (\mathcal{A}_r - \mu I)^{-k} \mathcal{B}_r K (A_n - \mu I)^q & -(\mathcal{A}_r - \mu I)^{-k} \mathcal{B}_r K (A_n - \mu I)^k & \mathcal{A}_r \end{bmatrix} \tag{3.130}$$

$$\mathcal{L} = \begin{bmatrix} 0 & \cdots & 0 & L \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \tag{3.131}$$

(see (3.80), (3.91), (3.93), (3.123) and (3.125)–(3.126)). In order to examine the spectrum of $\tilde{\mathcal{A}}_{cl,q}$ it is compared to the system operator $\mathcal{A}_{cl,q} = \mathcal{A}_{cl,0} + \Delta_q$ that was derived in Subsection 3.2.3 for the closed-loop system Σ_{cl}^{q} with respect to the untransformed states. This system operator is given by

$$\mathcal{A}_{cl,q} = \begin{bmatrix} A_n - LC_n & 0 & -L\mathcal{C}_r(\mathcal{A}_r - \mu I)^{-q} \\ B_n K (A_n - \mu I)^q & A_n - B_n K & 0 \\ \mathcal{B}_r K (\mathcal{A}_r - \mu I)^q & -\mathcal{B}_r K & \mathcal{A}_r \end{bmatrix} \tag{3.132}$$

(see (2.225), (3.95), and (3.98)) and apparently related to $\tilde{\mathcal{A}}_{cl,q}$ according to

$$\tilde{\mathcal{A}}_{cl,q} = \tilde{\mathcal{T}} \mathcal{A}_{cl,q} \tilde{\mathcal{T}}^{-1} \tag{3.133}$$

with

$$\tilde{\mathcal{T}} = \begin{bmatrix} I & 0 & 0 \\ 0 & (A_n - \mu I)^{-k} & 0 \\ 0 & 0 & (\mathcal{A}_r - \mu I)^{-k} \end{bmatrix}. \tag{3.134}$$

It is well-known that two linear operators that can be converted into one another by a linear transformation have coinciding spectra. However, $\tilde{\mathcal{T}}^{-1}$ is unbounded on X and hence not a transformation. In order to show that the spectra of $\mathcal{A}_{cl,q}$ and $\tilde{\mathcal{A}}_{cl,q}$ coincide nevertheless, the fact is used that the state $\tilde{x}_{cl}$ is confined to the subspace

$$\tilde{X}_{cl} := \{\tilde{\mathcal{T}}h,\, h \in X_{cl}\} \tag{3.135}$$

with $X_{cl} = \mathbb{C}^n \oplus \mathbb{C}^n \oplus X_r$ since

$$\tilde{x}_{cl} = \tilde{\mathcal{T}} x_{cl} \tag{3.136}$$

(see (3.94), (3.115), (3.122), (3.127) and (3.134)). Therefore, it is sufficient to consider the state space model $\tilde{\Sigma}^q_{cl}$ in (3.128)–(3.129) on the reduced state space $\tilde{X}_{cl}$. In this space the norm

$$\|h\|_{\tilde{X}_{cl}} := \|\tilde{\mathcal{T}}^{-1}h\|_{X_{cl}} \tag{3.137}$$

is used which is defined for all $h \in \tilde{X}_{cl}$ since for all $h = \tilde{\mathcal{T}}g$ with $g \in X_{cl}$ one has $\|h\|_{\tilde{X}_{cl}} = \|\tilde{\mathcal{T}}g\|_{\tilde{X}_{cl}} = \|g\|_{X_{cl}}$. This leads to the following result that is proven in Appendix A.10.

Theorem 3.3-1

The closed-loop system operator $\tilde{\mathcal{A}}_{cl,q}$ in (3.130) is the generator of a C_0-semigroup on the state space $\tilde{X}_{cl}$. Furthermore,

$$\sigma(\tilde{\mathcal{A}}_{cl,q}) = \sigma(\mathcal{A}_{cl,q}) \tag{3.138}$$

holds.

This statement shows that the spectrum of the closed-loop system is independent of the parameter k of the state transformations. The considerations in this section can be summarized therefore as follows: The compensator design method for reducing the spectrum perturbation was formulated in Section 3.2 such that only the observation spillover is reduced. However, whether the control spillover or the observation spillover or both are affected by the approach depends on the choice of the state variables for the approximation and the residual dynamics. Thus, the observation spillover reduction can be converted into control spillover reduction simply by transforming these states. The achieved suppression of the spectrum perturbation for a givcn q is in any case the same.

Chapter 4

The early-lumping approach for discrete-time control

In the previous chapters the compensator design for linear infinite-dimensional systems on the basis of the early-lumping approach was addressed, where the considerations referred to dynamical systems and controllers that operate in continuous time. While this assumption is appropriate for processes in many applications it deserves a more differentiated treatment regarding the compensators. These are implemented almost always by digital devices, which leads to a discrete-time operation.

In Figure 21 the structure of a sampled-data closed-loop system is shown. While a continuous-time operation is assumed for the infinite-dimensional plant Σ the observer-based compensator Σ_c^d works in discrete time at instances $t_k := kT$, $k \in \mathbb{N}_0$, of time, where $T > 0$ is the *sampling constant*. It generates the discrete-time control input $u_d[k]$, and is fed by the discrete-time measurement $y_d[k]$ as well as $u_d[k]$. For the conversion between the discrete-time signals u_d and y_d and the continuous-time signals u and y a *hold device* H and a *sampling device* S is used.

Often, the sampling constant is chosen small compared to the time constants of the controlled system. In this case, that is referred to as *quasi continuous-time control*, the sampled-data closed-loop system has almost the same behavior at the sampling time instances as the corresponding continuous-time control system. It is then admissible to apply the considerations in the Chapters 2–3 for continuous-time control although a discrete-time compensator is used. However, the smaller the sampling constant is, the more computational efforts are needed for evaluating the control algorithm. Therefore, it may be desirable to choose the sampling constant so large that a quasi continuous-

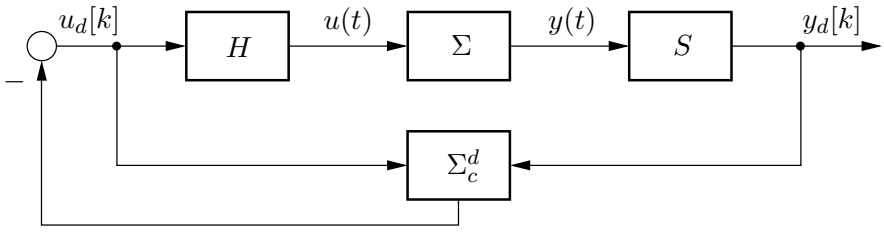

Figure 21 – Structure of a sampled-data control system consisting of a continuous-time plant Σ, a discrete-time observer-based compensator Σ_c^d, as well as a hold device H and a sampling device S. The latter ones convert the discrete-time control input $u_d[k]$, $k \in \mathbb{N}_0$, to the continuous-time plant input $u(t)$, $t \geq 0$, and the continuous-time plant output $y(t)$, $t \geq 0$, to the discrete-time measurement $y_d[k]$, $k \in \mathbb{N}_0$, respectively.

time treatment is not suitable and the sampled-data nature of the control system has to be taken into account.

When using the continuous-time control, a new value can be assigned to the input $u(t)$ at each and every instant of time t. In the discrete-time case in contrast, the compensator provides a new control input $u_d[k]$ only at the discrete sampling time instances t_k. In this way, the digital control is significantly restricted. This observation can be described by the fact that the resulting input u is confined in the discrete-time case to a certain subspace of $L_q([0,\tau];\mathbb{R}^p)$, $\tau > 0$, $q \in \mathbb{N}$, while these whole function spaces are admissible for continuous-time control (see Subsection 2.1.2). That the excitation of the plant is restricted in this way has the consequence that a number of technical difficulties of the continuous-time domain appearing for infinite-dimensional systems are avoided in the discrete-time case. In particular, the stability of both the open-loop and the closed-loop system is determined by the corresponding spectra, which is not true in general for continuous-time systems (see Subsection 2.1.3). At the same time, however, considering the system dynamics only at discrete instances of time leads to the shortcoming that the behavior of the plant within the sampling intervals is not covered by the system description so that its analysis requires additional efforts. Furthermore, inserting the hold and sampling devices into the control loop leads to the problem that additional controllability and observability defects may be introduced.

As can be expected, the compensator design for sampled-data systems on the basis of the early-lumping approach equals the procedure in the continuous-time case to a large extent. After an approximation of the infinite-dimensional plant has been de-

termined it is transformed to the corresponding sampled-data system. Subsequently, the compensator can be computed by the established approaches for finite-dimensional sampled-data systems (see, *e.g.*, [93]), whereat observer-based compensators are considered in this chapter as has been done before. It will be shown that the spillover discussed in Chapter 2 appears in an analog manner also for sampled-data systems which affects the closed-loop spectrum and thus influences the performance of the control loop (see also [9]). In order to analyze the spectrum perturbation the residual dynamics have to be taken into account for which purpose a discrete-time description of these continuous-time dynamics have to be determined. Then, the spectrum perturbation can be estimated analog to the statements in Section 2.4 for the continuous-time control.

The chapter is organized as follows. In Section 4.1 the representation of sampled-data systems is reviewed and the differences of their properties in comparison to continuous-time infinite-dimensional models are addressed. Furthermore, the class of systems in consideration is defined in that section. In Section 4.2 the compensator design by means of the early-lumping approach for discrete-time control is summarized and the spillover is characterized by analyzing the spectrum perturbation.

4.1 Discrete-time state space system representations

It comes from the nature of a digitally implemented compensator that its state $\hat{x}_n$ changes only at the discrete sampling time instances t_k, $k \in \mathbb{N}_0$, which suggests itself to describe it as a sequence $\hat{x}_n = (\hat{x}_{n,k})_{k\in\mathbb{N}_0}$ instead of a function that depends continuously on the time. Since the elements of this sequence correspond each to a certain instant of time t_k, such a sequence can be regarded as a map that is defined on the discrete set

$$\Xi := \{t_k = kT, k \in \mathbb{N}_0\}, \tag{4.1}$$

where $T > 0$ is the *sampling constant*. For instance, the compensator state $\hat{x}_n = (\hat{x}_{n,k})_{k\in\mathbb{N}_0}$ defines the map

$$\hat{x}_n : t \mapsto \hat{x}_{n,t/T}, \qquad \forall\, t \in \Xi. \tag{4.2}$$

In order to emphasize the character of a map depending on the time, any discrete-time signal $(\chi_k)_{k\in\mathbb{N}_0}$ will be referred to in the following by $\chi[k], k \in \mathbb{N}_0$. This notation is convenient because it is closer related to the notation of a continuous-time signal

$\nu(t), t \geq 0$, than the sequence form, where the square brackets indicate that the signal belongs to the discrete-time domain. Thus, $u_d[k]$ and $y_d[k]$ denote the discrete-time inputs and the output of the compensator (see Figure 21).

The hold device considered in this chapter is the commonly used *zero-order hold*. It keeps the last control instant $u_d[k]$ of the compensator during the current *sampling interval* $[t_k, t_{k+1})$ and passes it to the system input u, *i.e.*,

$$u(t) = u_d[k], \qquad t \in [t_k, t_{k+1}),\ k \in \mathbb{N}_0. \tag{4.3}$$

The *standard sampling device* simply takes the plant output y at the sampling time instances and passes it to the compensator, *i.e.*

$$y_d[k] = y(t_k), \qquad k \in \mathbb{N}_0. \tag{4.4}$$

For doing so it is important to note that y is continuous in general (see [46, Lem. 3.1.5]) so that $y(t_k)$ is well-defined for all $k \in \mathbb{N}_0$. Due to the output equation $y(t) = \mathcal{C}x(t)$ of the continuous-time system Σ (see (2.4)) this makes apparent that the compensator takes only the state x of Σ at the sampling time instances t_k into account while $x(t)$ for $t \notin \Xi$ does not influence the compensator at all. This motivates to describe the infinite-dimensional system Σ only at the sampling time instances. Thus, instead of using the model (2.3)–(2.4) a discrete-time model will be applied throughout this and the following chapter. Such a model is determined next.

4.1.1 State space models of sampled-data systems

The following considerations are based on the continuous-time plant model

$$\Sigma: \quad \dot{x}(t) = \mathcal{A}x(t) + \mathcal{B}u(t), \qquad t > 0, \quad x(0) = x_0 \in X \tag{4.5}$$

$$y(t) = \mathcal{C}x(t), \qquad t \geq 0 \tag{4.6}$$

with $u(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^m$, and $x(t) \in X$. Thereby, it is assumed as before that this model is a state linear system. That means that $\mathcal{B}$ and $\mathcal{C}$ are bounded and hence satisfy Assumption 2.1-2, and $\mathcal{A}$ is the generator of a C_0-semigroup (compare to Subsection 2.1.1). It is the aim of this subsection to determine a model that takes the hold device and the sampling device into account and that describes the system state x on the discrete-time axis. That means that the state

$$x[k] \coloneqq x(kT), \qquad k \in \mathbb{N}_0 \tag{4.7}$$

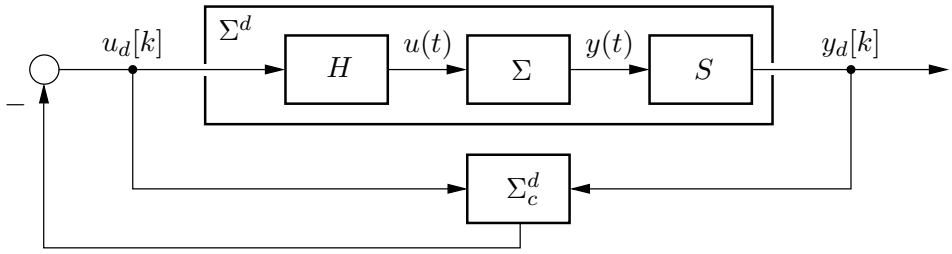

Figure 22 – The sampled-data system, that consists of the continuous-time plant Σ, the hold device H, and the sampling device S, is described by the infinite-dimensional discrete-time model Σ^d. It is controlled by the discrete-time observer-based compensator Σ_c^d.

is described by a discrete-time system Σ^d that models the aggregation of the continuous-time plant Σ as well as the hold device H and the sampling device S (see Figure 22). Such a combination of a continuous-time system with a hold and sampling device is referred to as *sampled-data system*. Note, that for both the continuous-time and the discrete-time signal the same identifier x is used for simplification of the notation. For the output y_d of this system (see Figure 22) it follows directly from (4.4) and (4.6) that

$$y_d[k] = y(t_k) = \mathcal{C}x(t_k) = \mathcal{C}x[k] \tag{4.8}$$

holds which is the output equation of Σ^d. In order to obtain a relation between $x[k] = x(t_k)$ and $x[k+1] = x(t_{k+1})$ the state equation (4.5) is solved formally using

$$x(t_{k+1}) = \mathcal{S}(t_{k+1} - t_k)x(t_k) + \int_{t_k}^{t_{k+1}} \mathcal{S}(t_{k+1} - \tau)\mathcal{B}u(\tau)\mathrm{d}\tau \tag{4.9}$$

(see [46, Thm. 3.1.7]), wherein $\mathcal{S}(t)$ is the C_0-semigroup generated by $\mathcal{A}$. Making use of the fact that the input u is constant within the sampling intervals due to (4.3) this can be transformed to

$$x[k+1] = \mathcal{S}(T)x[k] + \int_0^T \mathcal{S}(T - \tau)\mathrm{d}\tau\, \mathcal{B}u_d[k] \tag{4.10}$$

by substituting τ by $\tau - t_k$. Thus, when the operators

$$\mathcal{A}_d h := \mathcal{S}(T)h, \qquad \forall\, h \in X \tag{4.11}$$

$$\mathcal{B}_d v := \int_0^T \mathcal{S}(T - \tau)\mathrm{d}\tau\, \mathcal{B}v = \int_0^T \mathcal{S}(\tau)\mathrm{d}\tau\, \mathcal{B}v, \qquad \forall\, v \in \mathbb{C}^p \tag{4.12}$$

are introduced, where the right equation in (4.12) is obtained by aid of the substitution $\tau \to T - \tau$, one obtains from (4.8) and (4.10) the discrete-time model

$$\Sigma^d: \quad x[k+1] = \mathcal{A}_d x[k] + \mathcal{B}_d u_d[k], \qquad k \in \mathbb{N}_0, \quad x[0] = x_0 \in X \tag{4.13}$$

$$y_d[k] = \mathcal{C}x[k], \qquad k \in \mathbb{N}_0 \tag{4.14}$$

of the sampled-data system. Note, that (4.12) can be simplified to

$$\mathcal{B}_d v = \mathcal{A}^{-1}(\mathcal{S}(T) - I)\mathcal{B}v, \qquad \forall\, v \in \mathbb{C}^p \tag{4.15}$$

in case that $\mathcal{A}$ is invertible on X (see [46, Thm. 2.1.10]). An explicit expression for $\mathcal{A}_d$ and $\mathcal{B}_d$ can be obtained when $\mathcal{A}$ is a Riesz-spectral operator. Then, taking (2.26) and (2.14) into account for (4.11)–(4.12) yields

$$\mathcal{A}_d h = \sum_{i=1}^{\infty} e^{\lambda_i T} \langle h, \psi_i \rangle_X \phi_i, \qquad \forall\, h \in X \tag{4.16}$$

$$\mathcal{B}_d v = \sum_{i=1}^{\infty} \sum_{j=1}^{p} \int_0^T e^{\lambda_i \tau} \mathrm{d}\tau \, \langle b_j, \psi_i \rangle_X v_j \phi_i, \qquad \forall\, v \in \mathbb{C}^p, \tag{4.17}$$

wherein λ_i are the eigenvalues of $\mathcal{A}$, and ϕ_i and ψ_i denote the corresponding eigenvectors of $\mathcal{A}$ and $\mathcal{A}^*$, respectively. Note, that $\int_0^T e^{\lambda_i \tau} \mathrm{d}\tau$ in (4.17) can be simplified to

$$\int_0^T e^{\lambda_i \tau} \mathrm{d}\tau = \begin{cases} \frac{1}{\lambda_i}(e^{\lambda_i T} - 1) & : \quad \lambda_i \neq 0 \\ T & : \quad \lambda_i = 0. \end{cases} \tag{4.18}$$

It is shown later that $\mathcal{A}_d$ again is a Riesz-spectral operator if $\mathcal{A}$ is a Riesz-spectral operator that satisfies a mild condition (see Proposition 4.1-3). Comparison of (4.16) with the modal decomposition (2.33) of $\mathcal{A}$ shows that $\mathcal{A}_d$ has the same eigenvectors ϕ_i as $\mathcal{A}$, and its eigenvalues are

$$\lambda_{d,i} := e^{\lambda_i T}, \qquad i \in \mathbb{N}. \tag{4.19}$$

This shows that the eigenvalues $\lambda_{d,i}$ of $\mathcal{A}_d$ accumulate at $\lambda_d^{acc} = 0$ if $\operatorname{Re} \lambda_i \to -\infty$ for $i \to \infty$.

Of course, the model Σ^d provides a description of the system state only at the time instances $t \in \Xi$ so that no information is available about $x(t)$ for $t \notin \Xi$ directly from this model. However, restricting the class of input signals and confining the considerations to the discrete-time axis has the consequence that some basic properties, such as the solvability of the state equation (4.13), become simplified compared to the continuous-time counterpart. This is discussed in the following subsection.

4.1.2 Solution of the state equation and power stability

The difference equation and the initial condition (4.13) determine the state x on the discrete set Ξ of time instances t_k uniquely. This can be seen by applying (4.13)

recursively which directly yields the solution

$$x[k] = \mathcal{A}_d^k x_0 + \sum_{l=0}^{k-1} \mathcal{A}_d^{k-1-l} \mathcal{B}_d u_d[l], \qquad k \in \mathbb{N}, \quad \forall\, x_0 \in X,\, \forall u_d[l] \in \mathbb{C}^p. \tag{4.20}$$

Note, that $\mathcal{A}_d = \mathcal{S}(T)$ is a bounded operator in view of (2.25), so that it is unproblematic to consider the powers of this operator. This is an important difference to the system operator $\mathcal{A}$ of the continuous-time system Σ that is unbounded. For systems with $\mathcal{A}$ being a Riesz-spectral operator this becomes

$$\begin{aligned} x[k] &= \sum_{i=1}^{\infty} e^{k\lambda_i T} \langle x_0, \psi_i \rangle_X \phi_i + \sum_{l=0}^{k-1} \sum_{i=1}^{\infty} \sum_{j=1}^{p} e^{(k-1-l)\lambda_i T} \int_0^T e^{\lambda_i \tau} \mathrm{d}\tau \, \langle b_j, \psi_i \rangle_X u_{d,j}[l] \phi_i \\ &= \sum_{i=1}^{\infty} \lambda_{d,i}^k \langle x_0, \psi_i \rangle_X \phi_i + \sum_{l=0}^{k-1} \sum_{i=1}^{\infty} \sum_{j=1}^{p} \lambda_{d,i}^{k-1-l} \int_0^T e^{\lambda_i \tau} \mathrm{d}\tau \, \langle b_j, \psi_i \rangle_X u_{d,j}[l] \phi_i, \end{aligned} \tag{4.21}$$

where (2.21), (4.16)–(4.17), and (4.19) have been used.

The stability of discrete-time systems is commonly characterized by the *power stability* of $\mathcal{A}_d$. It is defined by the existence of constants $C \geq 1$ and $0 < \gamma < 1$ such that

$$\|\mathcal{A}_d^k\| \leq C\gamma^k, \qquad \forall\, k \in \mathbb{N}_0 \tag{4.22}$$

(see [97, Def. 2]). This relation implies that $\|x(t_k)\|$ decays exponentially toward zero for $u \equiv 0$ because in this case $x[k] = \mathcal{A}_d^k x_0$, $k \in \mathbb{N}$, follows from (4.20) so that $\|x[k]\|_X \leq \|\mathcal{A}_d^k\| \|x_0\|_X \to 0$ for $k \to \infty$. Moreover, if and only if the discrete-time system Σ^d is power stabilized, the continuous-time system Σ is stabilized exponentially so that $\|x(t)\|_X \to 0$ for $t \to \infty$ (see [111, Prop. 2.1]). A sufficient and necessary condition for $\mathcal{A}_d$ being power stable is

$$r_{\mathcal{A}_d} := \sup_{\lambda_d \in \sigma(\mathcal{A}_d)} |\lambda_d| < 1 \tag{4.23}$$

(see [97, Lem. 1]) which is the generalized version of the well-known condition for finite-dimensional discrete-time systems. Therein, $r_{\mathcal{A}_d}$ denotes the *spectral radius* of $\mathcal{A}_d$. More generally, for the homogeneous state solution $x[k] = \mathcal{A}_d^k x_0$ the relation

$$\|x[k]\|_X \leq C \|x_0\|_X \, \tilde{\beta}_d^k, \qquad \forall\, \tilde{\beta}_d > \beta_d,\ \forall\, k \in \mathbb{N}_0 \tag{4.24}$$

with a constant $C \geq 1$ and

$$\beta_d = r_{\mathcal{A}_d} \tag{4.25}$$

can be shown[1]. Since x approaches zero for $\beta_d < 1$ with a certain minimum decay rate, β_d is referred to as the *stability margin* of the discrete-time system, which is then called *power β_d-stable*. Thus, in contrast to the continuous-time case, where the (exponential) stability is determined by the spectrum only if the spectrum determined growth assumption (SDGA) holds (see Subsection 2.1.3), it is possible for discrete-time systems in general to analyze the power stability with the help of the spectrum $\sigma(\mathcal{A}_d)$. This relies basically on the fact that $\mathcal{A}_d = \mathcal{S}(T)$ is a bounded operator (see (2.25)) which is an important difference to $\mathcal{A}$ which is unbounded.

Usually, only the spectrum $\sigma(\mathcal{A})$ is known while $\sigma(\mathcal{A}_d)$ is unknown initially. The relation between both spectra is characterized by the following statements.

Proposition 4.1-1

The spectra $\sigma(\mathcal{A})$ and $\sigma(\mathcal{A}_d)$ are related by

$$\{e^{\lambda T} \,|\, \lambda \in \sigma(\mathcal{A})\} \subseteq \sigma(\mathcal{A}_d). \tag{4.26}$$

More specifically, the relations

$$\{e^{\lambda T} \,|\, \lambda \in \sigma_p(\mathcal{A})\} \subseteq \sigma_p(\mathcal{A}_d) \subseteq \{e^{\lambda T} \,|\, \lambda \in \sigma_p(\mathcal{A})\} \cup 0 \tag{4.27}$$

$$\sigma_r(\mathcal{A}_d) \subseteq \{e^{\lambda T} \,|\, \lambda \in \sigma_r(\mathcal{A})\} \tag{4.28}$$

hold.

The proof follows immediately from [107, Chapter 2, Thm. 2.3–2.5]. Unfortunately, a relation for $\sigma_c(\mathcal{A}_d)$ analog to (4.28) is *not* assured in general. In fact, it may exist a $\lambda_d \in \sigma_c(\mathcal{A}_d)$ that is not contained in $\{e^{\lambda T} \,|\, \lambda \in \sigma(\mathcal{A})\}$. Consequently, (4.26) is in general an inclusion but not an equality relation. As a simple example consider the spectrum

$$\sigma(\mathcal{A}) = \sigma_p(\mathcal{A}) = \{-1/k \pm j2\pi k \,|\, k \in \mathbb{N}\} \tag{4.29}$$

consisting of isolated eigenvalues. Equation (4.26) yields for $T = 1$

$$\{e^{-1/k \pm j2\pi k} \,|\, k \in \mathbb{N}\} = \{e^{-1/k} \,|\, k \in \mathbb{N}\} \subseteq \sigma(\mathcal{A}_d), \tag{4.30}$$

[1] For proving this one considers the operator $\tilde{\mathcal{A}}_d := \tilde{\beta}_d^{-1}\mathcal{A}_d$ with $\tilde{\beta}_d > \beta_d = r_{\mathcal{A}_d}$, that is power stable in view of $r_{\tilde{\mathcal{A}}_d} = \tilde{\beta}_d^{-1} r_{\mathcal{A}_d} < 1$ and (4.23). Thus, it holds $\|\mathcal{A}_d^k\| = \|\tilde{\beta}_d^k \tilde{\mathcal{A}}_d^k\| \leq C\tilde{\beta}_d^k$, $\forall\, \tilde{\beta}_d > \beta_d$, by (4.22) yielding (4.24).

where the left hand-side accumulates at $\lambda_d^{acc} = 1$ while $\sigma(\mathcal{A})$ has no accumulation point. This is an additional spectral point in $\sigma_c(\mathcal{A}_d)$ that is not contained in $\{e^{\lambda T} \,|\, \lambda \in \sigma(\mathcal{A})\}$. In consequence, it is not possible to conclude the power stability of $\mathcal{A}_d$ alone from the spectrum $\sigma(\mathcal{A})$ in general. This, however, is not surprising because it is known that the stability of continuous-time systems is determined by the spectrum of its system operator only under the SDGA (see Subsection 2.1.3). If this assumption holds, $\sup_{\lambda\in\sigma(\mathcal{A})} \operatorname{Re}\lambda < 0$ implies exponential stability of $\mathcal{S}(t)$, *i.e.*, there exist constants $C \geq 1$, $\omega < 0$ such that $\|\mathcal{S}(t)\| \leq Ce^{\omega t}$, $\forall\, t \geq 0$ (see (2.25)). Thus,

$$\|\mathcal{A}_d^k\| = \|\mathcal{S}^k(T)\| = \|\mathcal{S}(kT)\| \leq Ce^{\omega kT} = C(e^{\omega T})^k, \qquad \forall\, k \in \mathbb{N} \tag{4.31}$$

is satisfied[2], which shows that also power stability of $\mathcal{A}_d$ is assured in this situation in view of (4.22) and $e^{\omega T} < 1$.

In the next subsection a class of continuous-time systems is defined that is considered for sampled-data control in the following with the aim to avoid some technical difficulties.

4.1.3 Definition of the considered system class

It has been shown in the previous subsection that it is not easy in general to describe the continuous spectrum $\sigma_c(\mathcal{A}_d)$ in terms of the spectrum $\sigma(\mathcal{A})$ of the continuous-time plant's system operator so that this spectrum is unknown in many applications. Since the continuous spectrum of $\mathcal{A}_d$, however, plays an important role for the analysis of the spectrum perturbation caused by spillover, the focus will be restricted for the discrete-time control to systems with $\mathcal{A}$ being a Riesz-spectral operator. Thus, Assumption 2.1-9, that is the basis for the considerations in the continuous-time domain, is replaced from now on by the following one.

Assumption 4.1-2 (Properties of $\mathcal{A}$ for discrete-time control)

The system operator $\mathcal{A}$ of the continuous-time model Σ is assumed to have the following properties:

1. $\mathcal{A}$ is the generator of a C_0-semigroup.

[2] The relation $\mathcal{S}^k(T) = \mathcal{S}(kT)$ follows from the semigroup property $\mathcal{S}(t_1)\mathcal{S}(t_2) = \mathcal{S}(t_1 + t_2)$, $\forall\, t_1, t_2 \geq 0$, that yields $\mathcal{S}^2(T) = \mathcal{S}(2T)$ for $t_1 = t_2 = T$. Repeated application gives $\mathcal{S}^k(T) = \mathcal{S}(kT)$, $k \in \mathbb{N}$.

2. $\mathcal{A}$ is a Riesz-spectral operator.

3. The eigenvalues λ_i, $i \in \mathbb{N}$, of $\mathcal{A}$ satisfy

$$\lambda_i - \lambda_k \neq j\frac{2l\pi}{T}, \qquad \forall\, l \in \mathbb{N}_0 \tag{4.32}$$

for all $i, k \in \mathbb{N}$. ◀

Note, that it is avoided by Item 3 that several eigenvalues of $\mathcal{A}$ correspond to the same eigenvalue of $\mathcal{A}_d$ according to (4.19). In this way it is assured that the eigenvalues $\lambda_{d,i}$, $i \in \mathbb{N}$, of $\mathcal{A}_d$ are simple. Consequently, an eigenvalue $\lambda_{d,i}$ is modal controllable and modal observable whenever the corresponding eigenvalue λ_i of $\mathcal{A}$ has this property. In addition to this assumption also Assumption 2.1-2, that requires the boundedness of $\mathcal{B}$ and $\mathcal{C}$, has to hold also in the discrete-time case. For the relation between $\sigma(\mathcal{A})$ and $\sigma(\mathcal{A}_d)$ one has for this system class the following characterization.

Proposition 4.1-3

Let Assumption 4.1-2 hold. Then, the associated system operator $\mathcal{A}_d$ of the sampled-data system Σ^d is a sectorial Riesz-spectral operator. For its spectrum the relations

$$\sigma_p(\mathcal{A}_d) = \{e^{\lambda T} \,|\, \lambda \in \sigma_p(\mathcal{A})\} \tag{4.33}$$

$$\sigma_c(\mathcal{A}_d) = \overline{\sigma_p(\mathcal{A}_d)} \setminus \sigma_p(\mathcal{A}_d) \supseteq \{e^{\lambda T} \,|\, \lambda \in \sigma_c(\mathcal{A})\} \tag{4.34}$$

$$\sigma_r(\mathcal{A}_d) = \sigma_r(\mathcal{A}) = \emptyset \tag{4.35}$$

$$\sigma(\mathcal{A}_d) = \overline{\sigma_p(\mathcal{A}_d)} \tag{4.36}$$

are satisfied.

For the proof see Appendix A.11. Besides having the characterization (4.33)–(4.36) the considered class of system operators $\mathcal{A}$ has the benefit that $\mathcal{A}_d$ and $\mathcal{B}_d$ have the modal representations (4.16)–(4.17), and also the state equation can be solved in the explicit form (4.21). Furthermore, considering the defined class enables to apply the statements in Section 2.4 concerning the spectrum analysis of continuous-time systems also to sampled-data systems. Before this will be done in the next section the sampled-data system corresponding to an Euler-Bernoulli beam is determined.

Example 4.1-4 (Euler-Bernoulli beam with Kelvin-Voigt damping, continued)

For an Euler-Bernoulli beam with Kelvin-Voigt damping, that was introduced in Example 2.1-15, a discrete-time model shall be determined that results from adding a zero-order hold and a standard sampling device. Compared to Example 2.1-15 the considerations are specialized here to the beam length $\ell = 1$. There, a continuous-time model Σ on the state space $X := L_2(0,1) \oplus L_2(0,1)$ with the inner product $\langle \left[\begin{smallmatrix} g_1 \\ g_2 \end{smallmatrix}\right], \left[\begin{smallmatrix} h_1 \\ h_2 \end{smallmatrix}\right] \rangle_X := \int_0^1 g_1(z)\overline{h_1(z)}\mathrm{d}z + \int_0^1 g_2(z)\overline{h_2(z)}\mathrm{d}z$ was found. It was verified that $\mathcal{A}$ is a Riesz-spectral operator with the eigenvalues

$$\lambda_{\pm i} = \left(-\delta(i\pi)^2 \pm \sqrt{\delta^2(i\pi)^4 - 1}\right)(i\pi)^2, \qquad i \in \mathbb{N}, \tag{4.37}$$

that have an accumulation point $\lambda^{acc} = -\frac{1}{2\delta}$ (see (2.68) and (2.70)). The corresponding eigenvectors of $\mathcal{A}$ are given by

$$\phi_i(z) = \sin(i\pi z)\begin{bmatrix} 1 \\ \frac{-\lambda_i}{(i\pi)^2} \end{bmatrix}, \qquad i \in \mathbb{N} \tag{4.38}$$

$$\phi_{-i}(z) = \sin(i\pi z)\begin{bmatrix} \frac{1}{(i\pi)^2} \\ \frac{-\lambda_{-i}}{(i\pi)^4} \end{bmatrix}, \qquad i \in \mathbb{N}, \tag{4.39}$$

and the related biorthonormal sequence is

$$\psi_i(z) = \sin(i\pi z)\frac{-2(i\pi)^4}{\overline{\lambda_i} - \overline{\lambda_{-i}}}\begin{bmatrix} \frac{\overline{\lambda_{-i}}}{(i\pi)^4} \\ \frac{1}{(i\pi)^2} \end{bmatrix}, \qquad i \in \mathbb{N} \tag{4.40}$$

$$\psi_{-i}(z) = \sin(i\pi z)\frac{2(i\pi)^4}{\overline{\lambda_i} - \overline{\lambda_{-i}}}\begin{bmatrix} \frac{\overline{\lambda_i}}{(i\pi)^2} \\ 1 \end{bmatrix}, \qquad i \in \mathbb{N} \tag{4.41}$$

(see (2.71)–(2.72), and (3.107)–(3.108)). By taking the zero-order hold and the standard sampling device with sampling constant T into account one obtains from (4.16)–(4.17) the operators

$$\mathcal{A}_d h = \sum_{i \in \mathbb{Z}\setminus\{0\}} e^{\lambda_i T}\langle h, \psi_i\rangle_X \phi_i, \qquad \forall\, h \in X \tag{4.42}$$

$$\mathcal{B}_d v = \sum_{i \in \mathbb{Z}\setminus\{0\}} \frac{1}{\lambda_i}(e^{\lambda_i T} - 1)\left\langle \begin{bmatrix} 0 \\ \frac{1}{\beta_2 - \beta_1} \cdot \mathbf{1}_{[\beta_1,\beta_2]}(\cdot) \end{bmatrix}, \psi_i \right\rangle_X \phi_i v, \qquad \forall\, v \in \mathbb{C} \tag{4.43}$$

of the sampled-data system Σ_d, wherein (2.56) and $b(z) = \frac{1}{\beta_2 - \beta_1} \cdot \mathbf{1}_{[\beta_1,\beta_2]}(z)$ have been used. The output operator $\mathcal{C}$ of Σ_d coincides with that of the continuous-time system Σ (see (4.6) and (4.14)) and thus is given by

$$\mathcal{C}h = \left\langle h, \begin{bmatrix} \tilde{c} \\ 0 \end{bmatrix} \right\rangle_X, \qquad \forall\, h \in X \tag{4.44}$$

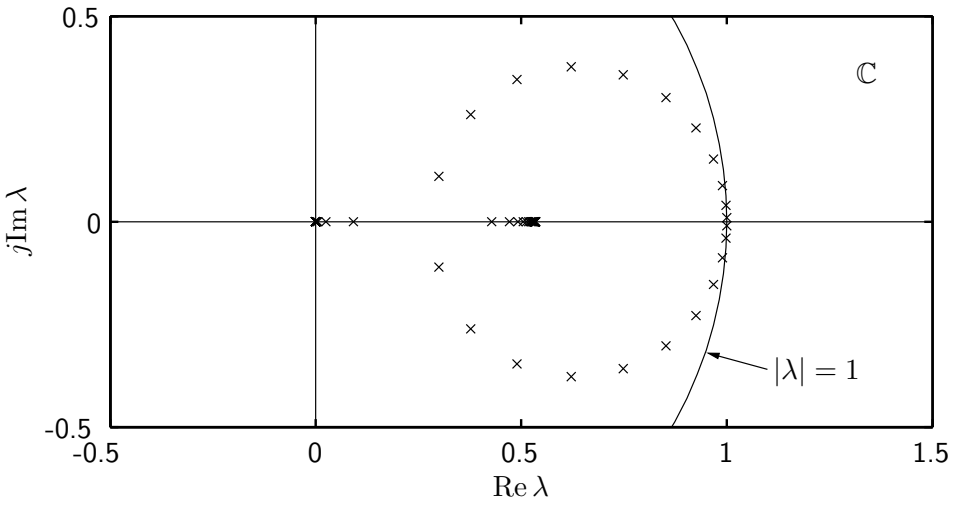

Figure 23 – Spectrum of the system operator $\mathcal{A}_d$ of the sampled-data system Σ^d that corresponds to the Euler-Bernoulli beam with structural damping for damping constant $\delta = 8 \cdot 10^{-4}$ and sampling constant $T = 10^{-3}$. While the system operator $\mathcal{A}$ of the continuous-time system has a single accumulation point $\lambda^{acc} = -625$, the spectrum of $\mathcal{A}_d$ has two accumulation points at $\lambda_{d,1}^{acc} = e^{\lambda^{acc}T} = 0.535$ and $\lambda_{d,2}^{acc} = 0$.

according to (2.58), in which $\tilde{c}(z)$ is defined in (2.60).

It has been shown in Example 2.1-15 that $\mathcal{A}$ satisfies Assumption 2.1-9 and hence is the generator of a C_0-semigroup. One can show easily that (4.32) holds for $T \neq l\pi/\mathrm{Im}\,\lambda_i, \forall\, l, i \in \mathbb{N}$. Since $\mathcal{A}$ is a Riesz-spectral operator that generates a C_0-semigroup, Assumption 4.1-2 is confirmed so that Proposition 4.1-3 can be applied. Therefore, $\mathcal{A}_d$ is a sectorial Riesz-spectral operator and its spectrum is given by

$$\sigma_p(\mathcal{A}_d) = \{e^{\lambda_{\pm i}T} \mid i \in \mathbb{N}\} \tag{4.45}$$

$$\sigma_c(\mathcal{A}_d) = \{0, e^{\lambda^{acc}T}\} \tag{4.46}$$

$$\sigma_r(\mathcal{A}_d) = \emptyset. \tag{4.47}$$

Note, that $\sigma(\mathcal{A}_d)$ has two accumulation points while $\sigma(\mathcal{A})$ has only one. For $\delta = 8 \cdot 10^{-4}$ and $T = 10^{-3}$ the eigenvalue distribution of $\mathcal{A}_d$ is shown in Figure 23. ◀

4.2 Observer-based control and analysis of the closed-loop spectrum

In this section the design of an observer-based compensator is addressed that is implemented digitally and thus operates in discrete time. Since the early-lumping approach is applied the design is based on an approximation that can be determined in two ways: In the first approach a discrete-time model Σ^d, that corresponds to the continuous-time plant Σ, is determined in a first step, and an approximation Σ_n^d of Σ^d is computed in a second one. In the second approach the continuous-time system Σ is reduced first yielding the continuous-time approximation Σ_n, which is converted subsequently to the discrete-time model Σ_n^d (see Figure 24). While both approaches lead to the same approximation Σ_n^d, the second one is certainly more relevant for the applications because both the approximation of the continuous-time system Σ and the time-discretization of the approximation can be computed by standard techniques. Therefore, this approach is considered here. To describe the approximation error, also the sampled-data system Σ_r^d corresponding to the continuous-time residual dynamics Σ_r is determined.

As in the continuous-time case the output feedback control design is considered without taking a reference input into account. However, this and the related feedforward control can be added separately by means of the two-degrees-of-freedom control method that is

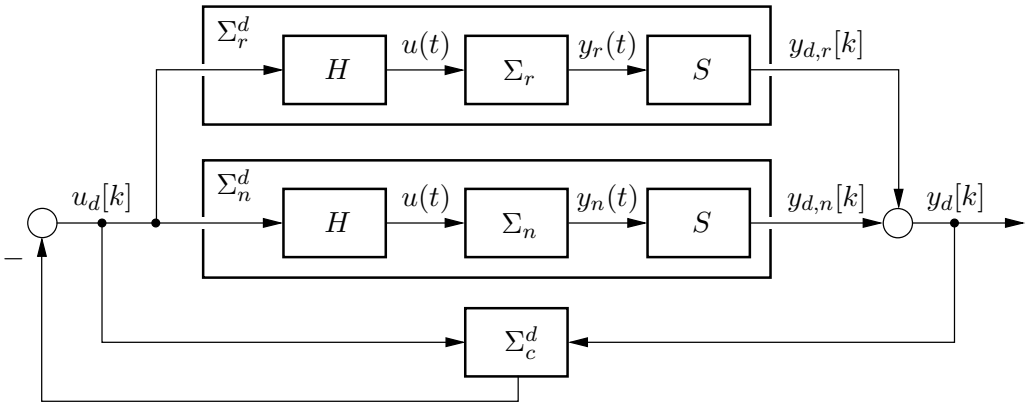

Figure 24 – Structure of a sampled-data control system. For applying the early-lumping approach the continuous-time infinite-dimensional plant is approximated by the finite-dimensional system Σ_n. In combination with the hold and sample devices H and S it yields the sampled-data approximation Σ_n^d. In the same way, the continuous-time residual dynamics Σ_r are converted to the sampled-data residual dynamics Σ_r^d.

mentioned at the beginning of Section 2.3. In addition, robust asymptotic disturbance rejection can be achieved by extending the continuous-time plant model by a suitable continuous-time signal model as it is explained in Subsection 2.3.4. However, one has to take care about the eigenvalues that are added in this way since Assumption 4.1-2 requires the system operator of the continuous-time system to be a Riesz-spectral operator which has simple eigenvalues (see Definition 2.1-6). For this reason, a more general class of Riesz-spectral operators, as defined in [73], has to be considered when, *e.g.*, a signal model for the rejection of ramp disturbances shall be used, which would have an eigenvalue $\lambda_s = 0$ with algebraic multiplicity $\nu_s = 2$. Such a generalization has to be considered also when the plant has an eigenvalue that coincides with one of the signal model.

In the following subsection the discrete-time systems Σ_n^d and Σ_r^d are determined. A digital observer-based compensator is designed in Subsection 4.2.2, and the resulting dynamics of the control loop is analyzed in Subsection 4.2.3.

4.2.1 Discrete-time system approximation

The compensator will be designed on the basis of the sampled-data system that corresponds to the modal approximation

$$\Sigma_n: \quad \dot{x}_n(t) = A_n x_n(t) + B_n u(t), \qquad t > 0, \quad x_n(0) = \mathcal{F}^{-1}\mathcal{P}x_0 \in \mathbb{C}^n \tag{4.48}$$

$$y_n(t) = C_n x_n(t), \qquad t \geq 0 \tag{4.49}$$

of order n, that was introduced in Section 2.2. For determining this system the hold device, whose operation is described by $u(t) = u_d[k]$ for $t \in [t_k, t_{k+1})$, $k \in \mathbb{N}_0$, and the sampling device yielding $y_{d,n}[k] = y_n(t_k) = y_n(kT)$ have to be taken into account. This can be carried out in the analog way as the conversion of the plant Σ to the sampled-data system Σ^d that was discussed in Subsection 4.1.1. The resulting discrete-time approximation is

$$\Sigma_n^d: \quad x_n[k+1] = A_{d,n} x_n[k] + B_{d,n} u_d[k], \quad k \in \mathbb{N}_0, \quad x_n[0] = \mathcal{F}^{-1}\mathcal{P}x_0 \in \mathbb{C}^n \tag{4.50}$$

$$y_{d,n}[k] = C_n x_n[k], \qquad k \in \mathbb{N}_0 \tag{4.51}$$

with

$$A_{d,n} = e^{A_n T} \tag{4.52}$$

$$B_{d,n} = \int_0^T e^{A_n \tau} \mathrm{d}\tau \; B_n \tag{4.53}$$

(see [93, Sec. 2.1]). In view of (2.108) one can evaluate (4.53) by

$$B_{d,n} = \operatorname{diag}\left(\int_0^T e^{\lambda_1 \tau}\mathrm{d}\tau, \ldots, \int_0^T e^{\lambda_n \tau}\mathrm{d}\tau\right) B_n \tag{4.54}$$

with

$$\int_0^T e^{\lambda_i \tau}\mathrm{d}\tau = \begin{cases} \frac{1}{\lambda_i}(e^{\lambda_i T} - 1) & : \quad \lambda_i \neq 0 \\ T & : \quad \lambda_i = 0. \end{cases} \tag{4.55}$$

By construction, the discrete-time state and the output of this model coincide with the respective state and output of the continuous-time approximation Σ_n at the sampling time instances, *i.e.*, $x_n[k] = x_n(t_k)$ and $y_{d,n}[k] = y_n(t_k)$, $k \in \mathbb{N}_0$. In order to describe the approximation error, the residual dynamics

$$\Sigma_r: \quad \dot{x}_r(t) = \mathcal{A}_r x_r(t) + \mathcal{B}_r u(t), \qquad t > 0, \quad x_r(0) = (I - \mathcal{P})x_0 \in X_r \tag{4.56}$$

$$y_r(t) = \mathcal{C}_r x_r(t), \qquad t \geq 0, \tag{4.57}$$

that was introduced in Section 2.2, is time-discretized yielding

$$\Sigma_r^d: \quad x_r[k] = \mathcal{A}_{d,r} x_r[k] + \mathcal{B}_{d,r} u_d[k], \qquad k \in \mathbb{N}_0, \quad x_r[0] = (I - \mathcal{P})x_0 \in X_r \tag{4.58}$$

$$y_{d,r}[k] = \mathcal{C}_r x_r[k], \qquad k \in \mathbb{N}_0 \tag{4.59}$$

with

$$\mathcal{A}_{d,r} h = \mathcal{S}_r(T)h, \qquad \forall\, h \in X_r \tag{4.60}$$

$$\mathcal{B}_{d,r} v = \int_0^T \mathcal{S}_r(\tau)\mathrm{d}\tau\, \mathcal{B}v, \qquad \forall\, v \in \mathbb{C}^p. \tag{4.61}$$

Thereby, $\mathcal{S}_r(t)$ is the C_0-semigroup generated by $\mathcal{A}_r$. Since $\mathcal{A}$ is a Riesz-spectral operator due to Assumption 4.1-2 one has the representation

$$\mathcal{S}_r(t)h = \sum_{i=n+1}^{\infty} e^{\lambda_i t}\langle h, \psi_i\rangle_X \phi_i, \qquad \forall\, h \in X_r, \tag{4.62}$$

which follows from (2.26) when restricted to X_r in view of the Riesz basis property. The derivation of Σ_r^d is analog to the approach for determining Σ^d in Subsection 4.1.1. In view of (4.60) and (4.62) $\mathcal{A}_{d,r}$ is bounded so that the difference equation (4.58) has a unique solution (see Subsection 4.1.2). Note, that $y_{d,r}[k] = y_r(t_k)$ and $x_r[k] = x_r(t_k)$, $k \in \mathbb{N}_0$, hold due to the construction of Σ_r^d. In view of

$$y_{d,n}[k] = C_n x_n[k] = C_n x_n(t_k) = y_n(t_k) \tag{4.63}$$

$$y_{d,r}[k] = \mathcal{C}_r x_r[k] = \mathcal{C}_r x_r(t_k) = y_r(t_k) \tag{4.64}$$

$$y_d[k] = \mathcal{C} x[k] = \mathcal{C} x(t_k) = y(t_k) \tag{4.65}$$

(see (4.6), (4.14), (4.49), (4.51), (4.57), and (4.59)) as well as $y(t) = y_n(t) + y_r(t)$ (see (2.120)) it holds

$$y_{d,n}[k] + y_{d,r}[k] = y_n(t_k) + y_r(t_k) = y(t_k) = y_d[k], \qquad \forall\, k \in \mathbb{N}_0, \tag{4.66}$$

which shows that the sampled-data models Σ_n^d and Σ_r^d of the approximation and the residual dynamics are complementary w.r.t. their outputs.

In general, controllability and observability may be lost when a hold and a sampling device is added to the closed-loop system for certain isolated values of the sampling constant $T > 0$. Necessary and sufficient conditions for $(A_{d,n}, B_{d,n})$ to be stabilizable are provided in [115]. Of course, it is assumed for the following, that $(A_{d,n}, B_{d,n})$ is controllable and $(C_n, A_{d,n})$ is observable, since this enables to assign the dynamics of the controlled system Σ_n^d by output feedback arbitrarily. These properties have to be checked after the approximation Σ_n^d has been determined. In case that a controllability or an observability defect has occurred in $(A_{d,n}, B_{d,n})$ or $(C_n, A_{d,n})$ that is not present in (A_n, B_n) and (C_n, A_n), respectively, a slight change of the sampling constant is typically sufficient to avoid the defect.

4.2.2 Design of the discrete-time observer-based compensator

In accordance with the explanations in Subsection 4.1.2 the discrete-time closed-loop system is power stable if and only if the spectrum of the corresponding system operator $\mathcal{A}_{d,cl}$ is entirely contained within the unit circle, *i.e.*,

$$\sup_{\lambda_d \in \sigma(\mathcal{A}_{d,cl})} |\lambda_d| < 1. \tag{4.67}$$

For stabilization of the control loop it is therefore necessary to shift all spectral points of the system operator $\mathcal{A}_d$, that are located outside the unit circle, into its interior by means of control. Instead of stability of the closed-loop system the stronger design objective is often desired that the closed-loop system is power β_d-stable for which purpose the spectral points of the system operator $\mathcal{A}_d$ located in

$$\mathbb{C}_{\beta_d}^o := \{s \in \mathbb{C} \;\big|\; |s| > \beta_d\} \tag{4.68}$$

have to be shifted into

$$\overline{\mathbb{C}}_{\beta_d}^i := \{s \in \mathbb{C} \;\big|\; |s| \leq \beta_d\} \tag{4.69}$$

(compare to (4.25)). According to Assumption 4.1-2 the system operators $\mathcal{A}$ and thus also $\mathcal{A}_d$ are Riesz-spectral operators for which reason the spectrum $\sigma(\mathcal{A}_d)$ consists of isolated eigenvalues $\lambda_{d,i}$ and possibly accumulation points (see Proposition 4.1-3). Analog to the continuous-time case the finite rank and the boundedness of both the input operator $\mathcal{B}_d$ and the output operator $\mathcal{C}$ allow to shift only a finite number[3] of eigenvalues from $\mathbb{C}^o_{\beta_d}$ into $\overline{\mathbb{C}}^i_{\beta_d}$. Therefore, the number of eigenvalues $\lambda_{d,i}$ of $\mathcal{A}_d$ that are contained in $\mathbb{C}^o_{\beta_d}$ must be finite. Particularly, they may not have any accumulation point in this set. Of course, the eigenvalues to be shifted by the control must be contained in the approximation in order to take them into account in the design procedure. These considerations lead to the following assumption, in which it is useful to remember that the eigenvalues λ_i of the continuous-time models Σ_n and Σ_r and the eigenvalues $\lambda_{d,i}$ of the discrete-time counterparts are related by $\lambda_{d,i} = e^{\lambda_i T}$, $i \in \mathbb{N}$, in view of (4.33).

Assumption 4.2-1 (Stabilizability of the closed-loop sampled-data system)

For the discrete-time systems Σ^d_n and Σ^d_r the following is assumed:

1. There are not more than finitely many eigenvalues of $\mathcal{A}_{d,n}$ located in $\mathbb{C}^o_{\beta_d} = \{s \in \mathbb{C} \mid |s| > \beta_d\}$, where $0 \leq \beta_d < 1$ is the desired stability margin of the discrete-time closed-loop system.
2. $(A_{d,n}, B_{d,n})$ is controllable and $(C_n, A_{d,n})$ is observable.
3. All spectral points of $\mathcal{A}_{d,r}$ are located in $\overline{\mathbb{C}}^i_{\beta_d} = \{s \in \mathbb{C} \mid |s| \leq \beta_d\}$. ◀

For distinct values of T Item 2 of this assumption may be not satisfied, although the continuous-time counterparts (A_n, B_n) and (C_n, A_n) are controllable and observable, respectively, as said in the previous subsection. Besides this aspect for choosing the sampling constant, T should be not larger than the largest time constant $T_{cl,i} := 1/|\lambda_{cl,i}|$ of the desired closed-loop eigenvalues $\lambda_{cl,i}$ in the continuous-time domain. Otherwise, the rejection of initial errors or (unmodeled) disturbance influences is limited by a too low sampling rate instead of the assigned dynamics. Thus, an unsatisfying *inter sampling behavior* would be the consequence (see also [61, Sec. 12.2]). Additionally, unstable modes of the approximation lead to large input amplitudes when the sampling

[3] This follows from [46, Thm. 5.2.6 and Thm. 5.2.7] when applied to sampled-data systems under use of (4.26).

constant is large. The discrete-time observer-based compensator to be designed has the form

$$\Sigma_c^d: \quad \hat{x}_n[k+1] = (A_{d,n} - LC_n)\hat{x}_n[k] + B_{d,n}u_d[k] + Ly_d[k], \qquad k \in \mathbb{N}_0 \tag{4.70}$$

$$\hat{x}_n[0] = \hat{x}_{n,0} \in \mathbb{C}^n \tag{4.71}$$

$$u_d[k] = -K\hat{x}_n[k], \qquad k \in \mathbb{N}_0. \tag{4.72}$$

Following the early-lumping approach the controller gain K and the observer gain L are chosen such that applying this compensator to the approximation Σ_n^d assigns the desired eigenvalues. It is well-known that these are the eigenvalues of the matrices $A_{d,n}-B_{d,n}K$ and $A_{d,n} - LC_n$ (see [93, Sec. 5.1]). Appropriate gains K and L can be determined by the established design methods for finite-dimensional eigenvalue assignment.

However, the compensator Σ_c^d is actually not applied to the approximation but to the infinite-dimensional sampled-data system Σ^d. The discrete-time residual dynamics Σ_r^d consequently influences the closed-loop behavior, leading to the same spillover effect that was discussed in the Chapters 2–3 for continuous-time control. This impact is investigated next.

4.2.3 Analysis of the closed-loop dynamics

In this subsection the behavior of the controlled system is analyzed in the analog way as it has been discussed in the Sections 2.3–2.4 for continuous-time control by investigating the spectrum of the closed-loop dynamics. For determining these dynamics the extended state

$$x_{cl}[k] := \begin{bmatrix} e_n[k] \\ x_n[k] \\ x_r[k] \end{bmatrix} \tag{4.73}$$

with $e_n[k] := x_n[k] - \hat{x}_n[k]$ is considered. By taking the system equations of Σ_n^d, Σ_r^d, and Σ_c^d into account (see (4.50)–(4.51), (4.58)–(4.59), and (4.70)–(4.72)), some basic manipulations result the closed-loop model

$$\Sigma_{cl}^d: \quad x_{cl}[k+1] = \mathcal{A}_{d,cl}x_{cl}[k], \qquad k \in \mathbb{N}_0, \quad x_{cl}[0] = x_{cl,0} \in X_{cl} \tag{4.74}$$

on the state space $X_{cl} = \mathbb{C}^n \oplus \mathbb{C}^n \oplus X_r$, wherein the closed-loop system operator is given by

$$\mathcal{A}_{d,cl} = \begin{bmatrix} A_{d,n} - LC_n & 0 & -L\mathcal{C}_r \\ B_{d,n}K & A_{d,n} - B_{d,n}K & 0 \\ \mathcal{B}_{d,r}K & -\mathcal{B}_{d,r}K & \mathcal{A}_{d,r} \end{bmatrix} \tag{4.75}$$

with $D(\mathcal{A}_{d,cl}) = X_{cl}$. Since all sub-blocks in $\mathcal{A}_{d,cl}$ are bounded this system operator is also bounded, which is why the discrete-time closed-loop system Σ_{cl}^d has a unique solution (compare to Subsection 4.1.2). Furthermore, the spectrum of $\mathcal{A}_{d,cl}$ determines the stability margin and thus the behavior of the control loop in the usual way as for finite-dimensional systems (see (4.24)–(4.25)). Particularly, the continuous-time plant Σ is stabilized in the sense that the norm $\|x(t)\|_X$ of its state decays exponentially if and only if $\mathcal{A}_{d,cl}$ is power stable. This can be shown in the same way as [111, Prop. 2.1].

For the spillover analysis the system operator $\mathcal{A}_{d,cl}$ is decomposed as

$$\mathcal{A}_{d,cl} = \mathcal{A}_{d,cl,0} + \Delta \tag{4.76}$$

with

$$\mathcal{A}_{d,cl,0} = \begin{bmatrix} A_{d,n} - LC_n & 0 & 0 \\ B_{d,n}K & A_{d,n} - B_{d,n}K & 0 \\ \mathcal{B}_{d,r}K & -\mathcal{B}_{d,r}K & \mathcal{A}_{d,r} \end{bmatrix}, \qquad \Delta = \begin{bmatrix} 0 & 0 & -L\mathcal{C}_r \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{4.77}$$

where $D(\mathcal{A}_{d,cl,0}) = D(\Delta) = X_{cl}$. Since $\mathcal{A}_{d,cl,0}$ has a triangular structure its spectrum is given by

$$\sigma(\mathcal{A}_{d,cl,0}) = \sigma(A_{d,n} - B_{d,n}K) \cup \sigma(A_{d,n} - LC_n) \cup \sigma(\mathcal{A}_{d,r}). \tag{4.78}$$

This makes apparent that the compensator, which is designed such that $A_{d,n} - B_{d,n}K$ and $A_{d,n} - LC_n$ have the desired eigenvalues, assigns these eigenvalues also to $\mathcal{A}_{d,cl,0}$. However, the closed-loop dynamics are described by the perturbed operator $\mathcal{A}_{d,cl} = \mathcal{A}_{d,cl,0} + \Delta$ so that the controlled system has the intended behavior only if $\|\Delta\| = \|L\mathcal{C}_r\|$ is sufficiently small. In general, however, all closed-loop eigenvalues $\tilde{\lambda}_{cl,i} \in \sigma(\mathcal{A}_{d,cl})$, $i \in \mathbb{N}$, differ from the desired values $\lambda_{cl,i} \in \sigma(\mathcal{A}_{d,cl,0})$, $i \in \mathbb{N}$, due to the spillover impact so that the performance of the controlled system may be unsatisfying. In this concern it is interesting to note, that the summands in (4.17), that correspond to eigenvalues λ_i of $\mathcal{A}_i$ in the left half-plane with large absolute values, become small in view of (4.18). In relation to the spillover effect this has the consequence that the corresponding eigenmodes, that belong to the residual dynamics, are qualitatively less influenced

than in the continuous-time case which suggests that the spillover becomes reduced just by using sampled-data control. However, small perturbations of the eigenvalues $\lambda_{d,i}$ close to the origin in the discrete-time domain correspond to larger changes of the related eigenvalues $\lambda_i = T^{-1} \ln \lambda_{d,i}$ (see (4.19)), wherein $\ln(\cdot)$ is the complex logarithm. Thus, the spillover impact qualitatively remains the same for discrete-time and continuous-time control.

The deviations of the eigenvalues and the structure of the closed-loop spectrum can be characterized by help of the results of Section 2.4 that have been presented there for the continuous-time control. Theorem 2.4-3, which gives a description of the closed-loop spectrum's structure, requires that the system operator $\mathcal{A}$ satisfies Assumption 2.1-9. In fact, since $\mathcal{A}$ is a Riesz-spectral operator that generates a C_0-semigroup due to Assumption 4.1-2, the corresponding system operator $\mathcal{A}_d$ of the sampled-data system Σ^d satisfies Assumption 2.1-9[4]. Since, in addition, $\mathcal{A}_{d,cl,0}$ has the same structure as the closed-loop system operator $\mathcal{A}_{cl,0}$ for the continuous-time case, and the perturbation operator Δ is even the same (compare (2.177) with (4.77)), Theorem 2.4-3 can readily be applied.

Corollary 4.2-2

Let the Assumption 4.1-2 hold. Then, the spectrum $\sigma(\mathcal{A}_{d,cl})$ of the closed-loop system operator in (4.75) can be decomposed as

$$\sigma(\mathcal{A}_{d,cl}) = \{\tilde{\lambda}_{cl,i}, i \in \mathbb{N}\} \cup \sigma_c(\mathcal{A}_d), \tag{4.79}$$

where $\tilde{\lambda}_{cl,i}$, $i \in \mathbb{N}$, are eigenvalues of $\mathcal{A}_{d,cl}$ that have finite algebraic multiplicities and are isolated. Particularly,

$$\sigma_r(\mathcal{A}_{d,cl}) = \emptyset \tag{4.80}$$

$$\sigma(\mathcal{A}_{d,cl}) = \overline{\sigma_p(\mathcal{A}_{d,cl})} \tag{4.81}$$

holds.

Thus, the closed-loop spectrum consists simply of isolated eigenvalues and of the accumulation points in $\sigma(\mathcal{A}_d)$. Now, the deviations of the eigenvalues caused by the

[4] This is stated by Proposition 2.1-10 when applied to $\mathcal{A}_d$. This is possible since this operator is a sectorial Riesz-spectral operator according to Proposition 4.1-3.

spillover are estimated, for which purpose the spectrum perturbation is redefined as

$$d_d := \sup_{\tilde{\lambda}_{cl} \in \sigma(\mathcal{A}_{d,cl})} \inf_{\lambda_{cl} \in \sigma(\mathcal{A}_{d,cl,0})} |\tilde{\lambda}_{cl} - \lambda_{cl}| \tag{4.82}$$

with $\mathcal{A}_{d,cl,0}$ and $\mathcal{A}_{d,cl}$ according to (4.77) and (4.75), respectively. An estimate for the spectrum perturbation can be obtained directly by aid of Theorem 2.4-10. For the applicability of this theorem Assumption 2.1-9 has to hold for $\mathcal{A}_d$ which is satisfied as discovered before. In addition, a transformation $\mathcal{T}_{cl} : X_{cl} \mapsto X_{cl}$ must exist such that $\mathcal{T}_{cl}^{-1}\mathcal{A}_{d,cl,0}\mathcal{T}_{cl}$ is normal. Since $\mathcal{A}_d$ is a Riesz-spectral operator and $\mathcal{A}_{d,cl,0}$ has the same structure as the continuous-time closed-loop system operator $\mathcal{A}_{cl,0}$, Proposition 2.4-8 can be applied by replacing $\mathcal{A}$, $\mathcal{A}_r$, A_n, and B_n by $\mathcal{A}_d$, $\mathcal{A}_{d,r}$, $A_{d,n}$, and $B_{d,n}$, respectively. It states that a normalizing transformation $\mathcal{T}_{cl}$ exists if $A_{d,n} - B_{d,n}K$ and $A_{d,n} - LC_n$ have simple and mutually different eigenvalues that are not contained in $\sigma(\mathcal{A}_{d,r})$. Thus, Theorem 2.4-10 can then be applied where, in doing so, it is again essential that $\mathcal{A}_{d,cl,0}$ and Δ have the same structure for both continuous-time and discrete-time control (compare (2.177) with (4.77)). This leads immediately to the following estimate for d_d.

Corollary 4.2-3

Let the Assumption 4.1-2 hold and assume that the observer-based compensator Σ_c^d in (4.70)–(4.72) is designed such that the eigenvalues of $A_{d,n} - B_{d,n}K$ and $A_{d,n} - LC_n$ are simple, mutually different, and not contained in $\sigma(\mathcal{A}_{d,r})$. Then,

$$d_d \leq \|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\| \leq \|\mathcal{T}_{cl}^{-1}\|\|\mathcal{T}_{cl}\|\|\Delta\| \tag{4.83}$$

is an upper bound for the spectrum perturbation d_d, wherein $\mathcal{T}_{cl}$ denotes a transformation such that $\mathcal{T}_{cl}^{-1}\mathcal{A}_{d,cl,0}\mathcal{T}_{cl}$ is normal.

As in the continuous-time case the evaluation of $\|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\|$ is not straightforward because $\mathcal{T}_{cl}$ and Δ are operators on the infinite-dimensional state space X. In fact, $\mathcal{T}_{cl}$ has the same structure as in the continuous-time domain, and the perturbation Δ is even the same. As before, the algorithm in Appendix B can be applied for the evaluation of $\|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\|$ by means of standard numerical computing software, provided that $\mathcal{A}$ is normal. Otherwise this norm has to be determined on the basis of an accurate approximation of the closed-loop dynamics.

An alternate estimate for d_d, that will be used in Subsection 5.1.3, can be obtained by reformulating the closed-loop dynamics Σ_{cl}^d as

$$\tilde{\Sigma}_{cl}^d: \quad \tilde{x}_{cl}[k+1] = (\tilde{\mathcal{A}}_{d,cl,0} + \tilde{\Delta})\tilde{x}_{cl}[k], \qquad k \in \mathbb{N}_0, \quad \tilde{x}_{cl}[0] = \tilde{x}_{cl,0} \in X_{cl} \tag{4.84}$$

with

$$\tilde{x}_{cl}[k] := \begin{bmatrix} x_n[k] \\ e_n[k] \\ x_r[k] \end{bmatrix} \tag{4.85}$$

and

$$\tilde{\mathcal{A}}_{d,cl,0} = \begin{bmatrix} A_{d,n} - B_{d,n}K & B_{d,n}K & 0 \\ 0 & A_{d,n} - LC_n & -L\mathcal{C}_r \\ 0 & 0 & \mathcal{A}_{d,r} \end{bmatrix}, \quad \tilde{\Delta} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\mathcal{B}_{d,r}K & \mathcal{B}_{d,r}K & 0 \end{bmatrix}, \tag{4.86}$$

where $D(\tilde{\mathcal{A}}_{d,cl,0}) = D(\tilde{\Delta}) = X_{cl}$. This form of the closed-loop dynamics follows from (4.73)–(4.75) when the order of the states e_n and x_n is interchanged in the composed state vector (compare (4.73) and (4.85)). Since $\tilde{\mathcal{A}}_{d,cl,0}$ has, as $\mathcal{A}_{d,cl,0}$, the desired eigenvalues of $A_{d,n} - B_{d,n}K$, $A_{d,n} - LC_n$, and $\mathcal{A}_{d,r}$ due to its triangular structure, this shows that the control loop behaves as intended if $\|\mathcal{B}_{d,r}K\|$ and thus $\|\tilde{\Delta}\|$ is negligible. This observation is characterized quantitatively by the following statement.

Theorem 4.2-4

Let the Assumption 4.1-2 hold and assume that the observer-based compensator Σ_c^d in (4.70)–(4.72) is designed such that the eigenvalues of $A_{d,n} - B_{d,n}K$ and $A_{d,n} - LC_n$ are simple, mutually different, and not contained in $\sigma(\mathcal{A}_{d,r})$. Then,

$$d_d \leq \|\tilde{\mathcal{T}}_{cl}^{-1}\tilde{\Delta}\tilde{\mathcal{T}}_{cl}\| \leq \|\tilde{\mathcal{T}}_{cl}^{-1}\|\|\tilde{\mathcal{T}}_{cl}\|\|\tilde{\Delta}\| \tag{4.87}$$

is an upper bound for the spectrum perturbation d_d, where $\tilde{\mathcal{T}}_{cl}$ denotes a linear transformation such that $\tilde{\mathcal{T}}_{cl}^{-1}\tilde{\mathcal{A}}_{d,cl,0}\tilde{\mathcal{T}}_{cl}$ is normal.

Analog to Theorem 2.4-10 this result follows immediately from Lemma 2.4-6 when applied to $\tilde{\mathcal{A}}_{d,cl} := \tilde{\mathcal{A}}_{d,cl,0} + \tilde{\Delta}$. This is possible because $\tilde{\Delta}$ is a bounded perturbation operator, $\sigma_r(\tilde{\mathcal{A}}_{d,cl}) = \sigma_r(\mathcal{A}_{d,cl}) = \emptyset$ holds according to Corollary 4.2-2. Finally, the existence of a normalizing transformation is ensured by Assumption 4.1-2 as argued above, because the difference between the state vectors $\tilde{x}_{cl}$ and x_{cl} is only the order

of the first two vector elements. Thus, all requirements of Lemma 2.4-6 are met. Its application to $\tilde{\mathcal{A}}_{d,cl}$ leads to Theorem 4.2-4.

Remark 4.2-5 For achieving asymptotic disturbance rejection w.r.t. the discrete-time dynamics a signal model of the exogenous disturbance has to be incorporated into the control loop (see Subsection 2.3.4). The considerations in this and the following chapter remain true for the modified control structure when the matrices in (4.77) are replaced by

$$\mathcal{A}_{d,cl,0} := \begin{bmatrix} A_{d,n} - LC_n & 0 & 0 & 0 \\ 0 & A_{d,s} & B_{d,s}C_n & 0 \\ B_{d,n}K & -B_{d,n}K_s & A_{d,n} - B_{d,n}K & 0 \\ \mathcal{B}_{d,r}K & -\mathcal{B}_{d,r}K_s & -\mathcal{B}_{d,r}K & \mathcal{A}_{d,r} \end{bmatrix} \tag{4.88}$$

$$\Delta := \begin{bmatrix} 0 & 0 & 0 & -L\mathcal{C}_r \\ 0 & 0 & 0 & B_{d,s}\mathcal{C}_r \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{4.89}$$

and the matrices in (4.86) are replaced by

$$\tilde{\mathcal{A}}_{d,cl,0} := \begin{bmatrix} A_{d,s} & B_{d,s}C_n & 0 & B_{d,s}\mathcal{C}_r \\ -B_{d,n}K_s & A_{d,n} - B_{d,n}K & B_{d,n}K & 0 \\ 0 & 0 & A_{d,n} - LC_n & -L\mathcal{C}_r \\ 0 & 0 & 0 & \mathcal{A}_{d,r} \end{bmatrix} \tag{4.90}$$

$$\tilde{\Delta} := \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\mathcal{B}_{d,r}K_s & -\mathcal{B}_{d,r}K & \mathcal{B}_{d,r}K & 0 \end{bmatrix}. \tag{4.91}$$

Therein, $A_{d,s} = e^{A_s T}$ and $B_{d,s} = \int_0^T e^{A_s \tau} \mathrm{d}\tau B_s$ are the discrete-time counterparts of A_s and B_s, respectively (compare to (4.52)–(4.53)). ◀

The design of a discrete-time observer-based compensator and the analysis of the closed-loop spectrum are demonstrated in the following example.

Example 4.2-6 (Euler-Bernoulli beam with Kelvin-Voigt damping, continued)

The sampled-data system determined in Example 4.1-4 for an Euler-Bernoulli beam with length $\ell = 1$ and Kelvin-Voigt damping is considered here for designing a discrete-time compensator. The damping constant is chosen as $\delta = 0.005$ and the input

and output distribution functions are given by

$$b(z) = \frac{1}{2\varepsilon} \cdot \mathbf{1}_{[0.4-\varepsilon,0.4+\varepsilon]}(z), \qquad c(z) = \frac{1}{2\varepsilon} \cdot \mathbf{1}_{[0.6-\varepsilon,0.6+\varepsilon]}(z) \tag{4.92}$$

with $\varepsilon = 10^{-3}$. Some of the eigenvalues of the continuous-time system operator $\mathcal{A}$ with largest real parts, that are given by (4.37), are listed in Table 5. In order to shift the eigenvalues $\lambda_{\pm 1}$ and $\lambda_{\pm 2}$ more to the left in the complex for reducing the settling time of the system response an observer-based compensator is designed on the basis of a 4-th-order sampled-data modal approximation Σ_n^d. For determining Σ_n^d the continuous-time approximation Σ_n with order $n = 4$ of the continuous-time plant Σ has to be computed first which subsequently is converted to a sampled-data system. The approximation Σ_n is obtained from (2.108)–(2.110) when the eigenvectors ϕ_i and the vectors ψ_i of the corresponding biorthonormal sequence are inserted, which are given by (4.38)–(4.41). The resulting parameters A_n, B_n, and C_n of the continuous-time approximation lead, by help of (4.52)–(4.53), to the discrete-time approximation Σ_n^d with

$$A_{d,n} = \operatorname{diag}\big(0.9581 \pm j0.2410,\, 0.4669 \pm j0.6777\big) \tag{4.93}$$

$$B_{d,n} = \begin{bmatrix} -0.00289 + j0.02342 \\ -0.02857 - j0.23117 \\ -0.00591 + j0.01169 \\ -0.23314 - j0.46144 \end{bmatrix} \tag{4.94}$$

$$C_n = \begin{bmatrix} -0.09636 & -0.00976 & 0.01489 & 0.000377 \end{bmatrix}. \tag{4.95}$$

Taking $A_{d,n}$, $B_{d,n}$, and C_n into account, it can be verified easily that $(A_{d,n}, B_{d,n})$ is controllable and $(C_n, A_{d,n})$ is observable so that Item 2 of Assumption 4.2-1 holds.

Next, the observer-based compensator Σ_c^d (see (4.70)–(4.72)) is designed by eigenvalue assignment. Instead of choosing the eigenvalues $\lambda_{d,c,i} = \sigma(A_{d,n} - B_{d,n}K)$ of the controlled discrete-time approximation and the eigenvalues $\lambda_{d,o,i} = \sigma(A_{d,n} - LC_n)$ of the

Table 5 – Eigenvalues λ_i of the continuous-time system operator $\mathcal{A}$ with largest real parts and corresponding eigenvalues $\lambda_{d,i}$ of the discrete-time system operator $\mathcal{A}_d$.

i	$\lambda_{\pm i}$	$\lambda_{d,\pm i}$	$\lvert\lambda_{d,\pm i}\rvert$
1	$-0.487 \pm j9.86$	$0.9581 \pm j0.2410$	0.9879
2	$-7.79 \pm j38.70$	$0.4669 \pm j0.6777$	0.8229
3	$-39.45 \pm j79.59$	$-0.1517 \pm j0.3407$	0.3730
4	$-124.68 \pm j96.91$	$-0.0333 \pm j0.0292$	0.0443

discrete-time observer dynamics directly, it is easier to select the eigenvalues $\lambda_{c,i}$ and $\lambda_{o,i}$ of the corresponding continuous-time counterparts of these dynamics since the influence of the eigenvalues on the behavior of a continuous-time dynamics is well-known (see, *e.g.*, (2.24)). In order to reduce the settling time of the system response and to increase the damping the eigenvalues $\lambda_{\pm 1}$ and $\lambda_{\pm 2}$ are shifted by choosing $\lambda_{c,\pm 1} = -10 \pm j5$, $\lambda_{c,\pm 2} = -10 \pm j10$ and $\lambda_{o,\pm 1} = -15 \pm j5$, $\lambda_{o,\pm 2} = -15 \pm j10$. Thus, for assuring a stability margin $\beta = 7$ of the controlled system in continuous time the spectrum perturbation must satisfy $d \leq 3$ since the maximal real parts of the desired eigenvalues is given by $\operatorname{Re} \lambda_{c,i} = -10$. The sampling constant, is chosen as $T = 1/40 = 0.025$ so that (4.32) is satisfied. Thus, as claimed in Subsection 4.2.2, it is smaller than the largest time constant $T_{cl,i} = 1/|\lambda_{cl,i}|$ of the desired closed-loop eigenvalues $\{\lambda_{cl,i}, i \in \mathbb{N}\} = \{\lambda_{c,i}, i = \pm 1, \pm 2\} \cup \{\lambda_{o,i}, i = \pm 1, \pm 2\} \cup \sigma(\mathcal{A}_r)$ in the continuous-time domain (for $\sigma(\mathcal{A}_r)$ see Table 5). The corresponding eigenvalues to be assigned in the discrete-time domain are obtained from

$$\lambda_{d,c,i} = e^{\lambda_{c,i} T}, \qquad \lambda_{d,o,i} = e^{\lambda_{o,i} T}, \qquad i = \pm 1, \pm 2 \tag{4.96}$$

(compare to (4.19)), and the stability margin of the discrete-time closed-loop system is $\beta_d = e^{\beta T} = 0.839$. Since there are two and hence finitely many eigenvalues $\lambda_{d,i}$ located within $\mathbb{C}^o_{\beta_d}$ (see Table 5), and since these eigenvalues belong to the approximation Σ^d_n (see (4.93)), Assumption 4.2-1 is satisfied which enables to design a compensator that achieves the desired stability margin. Note, that the stability margin is limited by the accumulation points of $\sigma(\mathcal{A}_d)$ that are $\lambda^{acc}_{d,1} = 0.0821$ and $\lambda^{acc}_{d,2} = 0$ (see (4.46) and (2.70)).

In Example 4.1-4 it has been shown that Assumption 4.1-2 is satisfied so that Corollary 4.2-2 can be applied. It states that the spectrum $\sigma(\mathcal{A}_{d,cl})$ of the closed-loop system operator (4.75) consists of isolated eigenvalues $\tilde{\lambda}_{cl,i}$, $i \in \mathbb{N}$, with finite algebraic multiplicities and the accumulation points $\lambda^{acc}_{d,1}$ and $\lambda^{acc}_{d,2}$ of $\mathcal{A}_d$. The resulting eigenvalue distribution of the closed-loop system is depicted in Figure 25. It shows that five eigenvalues of the closed-loop system are located outside the region $\overline{\mathbb{C}}^i_{\beta_d}$ so that the specified control performance is not achieved, which reflects the impact of the spillover. The requirements of Corollary 4.2-3 are satisfied in view of the mentioned eigenvalue assignments and since Assumption 4.1-2 holds. Applying this corollary yields the bound $d_d \leq 15.7$ for the spectrum perturbation, where the computations have been carried out on the basis of a 60-th-order modal approximation. A computation of the eigenvalues by using this accurate approximation reveals that the perturbation is actually

$d_d = 0.329$, which confirms the correctness of the estimate, though being rather conservative. ◀

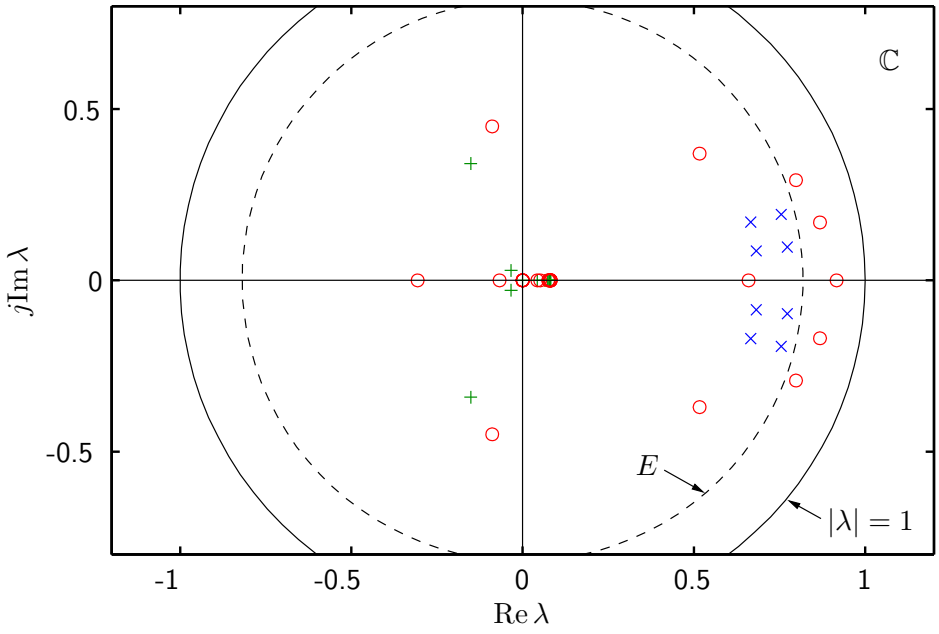

Figure 25 – Closed-loop eigenvalues ('o') in comparison with the desired eigenvalues. The desired ones consist of the eigenvalues assigned to $A_{d,n} - LC_n$ and $A_{d,n} - B_{d,n}K$ ('x') and those of $\mathcal{A}_{d,r}$ ('+'). Only some of the latter eigenvalues are visible in the plot because most of them are located close to the accumulation points $\lambda_{d,1}^{acc} = 0.0821$ and $\lambda_{d,2}^{acc} = 0$. The dashed circle defines the boundary of the region $\overline{\mathbb{C}}_{\beta_d}^{i}$ that corresponds to the specified control performance, and the circle with solid line describes the region of stability.

Chapter 5

Spillover reduction for discrete-time control

It was shown in the previous chapter that the early-lumping approach for the design of sampled-data control systems leads to the same spillover effect as in the continuous-time case that was discussed in Chapter 2. It is straightforward to adapt the spillover reduction approach of Chapter 3 to discrete-time control. To this end, the continuous-time output observers introduced in Section 3.1 have to be converted to the discrete-time version, which is straightforward in the light of Subsection 4.2.1. By extending the discrete-time plant Σ^d (see (4.13)–(4.14)) by a sufficient number of such output observers the spillover effect can be made arbitrarily small. The results of Subsection 3.2.3 about the reduction of the spectrum perturbation can be adapted to the discrete-time description of the closed-loop system. Thus, this spillover reduction approach is typically more efficient, as in the continuous-time case, than the conventional approach to increase simply the approximation order.

Alternatively to the approach based on output observers the control of sampled-data systems offers degrees of freedom that do not exist for continuous-time control, namely the choice of the hold device and the sampling device. In this chapter it is therefore the basic idea to use this freedom for spillover suppression. For that, a *general hold device* or a *general sampling device* is used instead of the zero-order hold and the standard sampling that was considered previously.

The *hold functions* $\theta_i(t)$ of a hold device, defined on the sampling interval $[0, T]$, describe the relation between an element $u_{d,i}[k]$ of the discrete-time control vector $u_d[k]$ and the corresponding continuous-time system input $u_i(t)$ during the current sampling

interval $[t_k, t_{k+1})$, *i.e.*,

$$u_i(t) = \theta_i(t - t_k)u_{d,i}[k], \qquad t \in [t_k, t_{k+1}),\ k \in \mathbb{N}_0,\ i = 1, 2, \ldots, p. \tag{5.1}$$

In contrast to a zero-order hold, where the hold functions are simply constant, a general hold device can have in principal arbitrary bounded hold functions. Since these determine the excitation of the continuous-time residual dynamics Σ_r, the freedom in the choice of the hold functions can be used to influence Σ_r in a desired way. Particularly, the hold functions will be chosen in the following such that the state of the continuous-time residual dynamics has a small norm at the sampling time instances t_k. Thus, when a standard sampling device is used that passes the measurements $y_d[k] = y(t_k)$ to the compensator the contributions of the residual dynamics do hardly appear in the discrete-time signal $y_d[k]$. In this way, the influence of these dynamics on the compensator becomes small so that the spillover effect is suppressed. The fact that the general hold device causes the state of the continuous-time residual dynamics to have a small norm at the sampling time instances has the meaning that the excitation of the corresponding sampled-data system becomes reduced. In other words, it is the aim of the general hold device to suppress the control spillover while the observation spillover is not affected (see Subsection 2.3.2).

The implementation of the general hold device becomes particularly simple when the hold functions are step functions, since it is then possible to realize the general hold by a conventional zero-order hold that operates at a higher sampling rate. This class of hold functions is therefore considered in the following. When the step heights are considered as the free parameters of a step function it is clear that only a finite number of degrees of freedom is available for reducing the spillover effect. For this reason, the spectrum perturbation caused by spillover cannot be eliminated entirely, but can be made arbitrarily small by choosing the number of steps sufficiently large. An estimate of the resulting spectrum will be provided. The approach requires to solve the state equations for both the approximation and the residual dynamics for a certain input signal. While the solution of the approximation model can be determined by standard software, the solution of the residual dynamics can be computed by a suitable PDE solver or on the basis of an accurate approximation.

The spillover reduction approach based on output observers and the alternative method involving a general hold device can be compared transparently when the continuous-time output $u(t)$ of the general hold is regarded as the response of a dynamical system that is excited by the discrete-time control $u_d[k]$. To be more precise, the general

hold has the effect of a *finite impulse response (FIR) filter* since the excitation $u_d[0] = [1\ 1\ \cdots\ 1]^T$ and $u_d[k] = 0$, $k \geq 1$, yields the signal $u(t) = [\theta_1(t)\ \theta_2(t)\ \cdots\ \theta_p(t)]^T$ on the finite interval $t \in [0, T]$ and $u(t) = 0$ for $t > T$. For this reason both spillover reduction methods introduce additional dynamics, which are contained in the output observers on the one hand, and in the general hold device on the other. The implementation of both types of dynamics, however, is apparently quite different. While a conventional discrete-time dynamical system is implemented as a recursion formula, the general hold device can be implemented by reading the step heights of the hold functions θ_i directly out of a memory and multiplying them by the control $u_{d,i}[k]$. For both approaches the order of the additional dynamics has to be increased for increasing the extent of spillover reduction. Which of both is more suitable for a specific application depends on several aspects. Particularly, the actor has to operate at a high sampling rate for realizing the general sampling law. In return, the output observer based approach leads to a comparatively complex structure of the control loop and may require more efforts for the design procedure.

Alternatively to the use of a general hold device it will be shown that the spillover can also be suppressed in a similar way by using a standard zero-order hold device but replacing the conventional sampling device by a general sampling device. It generates the instances $y_d[k]$ of the discrete-time output signal by computing the mean value of the weighted system output $y(t)$ over the previous sampling interval, *i.e.*,

$$y_{d,i}[k] = \int_{t_{k-1}}^{t_k} \pi_i(\tau - t_{k-1}) y_i(\tau) \mathrm{d}\tau, \qquad i = 1, 2, \ldots, m, \tag{5.2}$$

where $y_{d,i}$ and y_i denote the i-th element of the vectors y_d and y, respectively. The related weighting functions $\pi_i(t)$ are called *sampling functions*, and are chosen such that the contributions of the residual dynamics Σ_r to $y_d[k]$ are kept small. Consequently, the observation spillover is suppressed (see Subsection 2.3.2), in contrast to the approach with a general hold, where the control spillover becomes small. In general, it is not easy to implement the general sampling law since it involves an integration of the weighted continuous-time system output. In order to avoid this problem the sampling functions are restricted to a certain set of functions. The integration reduces then to a finite sum of output values that are obtained from standard sampling with higher sampling rate. Thus, the general sampling can be implemented in this way easily. The comparison of this approach to the output observer based spillover reduction technique is similar as discussed for the use of a general hold device. Also the general sampling device can be regarded as a dynamical system whose order increases for a larger extent of

spillover suppression. Different from the use of a general hold device not the actor but the sensor has to operate with a high sampling rate for general sampling, which may be easier to implement.

In the first section of this chapter the spillover reduction technique for sampled-data control systems with a general hold device and a standard sampling device is presented. The design of the hold functions is discussed and an estimate for the spectrum perturbation is given. Section 5.2 addresses the sampled-data control scheme with a zero-order hold device and a general sampling device, where again a design procedure and a perturbation estimate are presented.

5.1 Control using general hold devices

In this section the control of sampled-data systems is discussed that contain a general hold device while the sampling device is of the standard type (see Figure 26). The functionality of the considered general hold device H_Θ can be described by

$$u(t) = \Theta(t - t_k)u_d[k], \qquad t \in [t_k, t_{k+1}),\ k \in \mathbb{N}_0 \tag{5.3}$$

with

$$\Theta(t) = \operatorname{diag}\big(\theta_i(t), \ldots, \theta_p(t)\big). \tag{5.4}$$

Therein, the *hold functions*

$$\theta_i : [0, T] \mapsto \mathbb{R}, \qquad i = 1, 2, \ldots, p \tag{5.5}$$

are arbitrary but bounded, *i.e.*, $\theta_i \in L_\infty(0, T)$ (see Appendix D), since only bounded control inputs can be implemented, and p is the number of inputs, *i.e.*, the length of

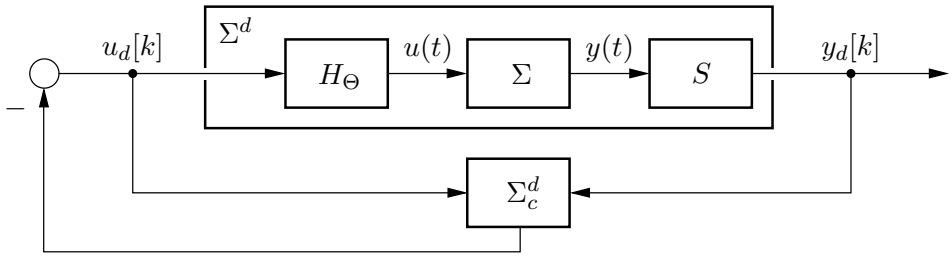

Figure 26 – Structure of a control system with a sampled-data system Σ^d and a discrete-time observer-based compensator Σ_c^d. The first one consists of a continuous-time plant Σ, a general hold device H_Θ, and a standard sampling device S.

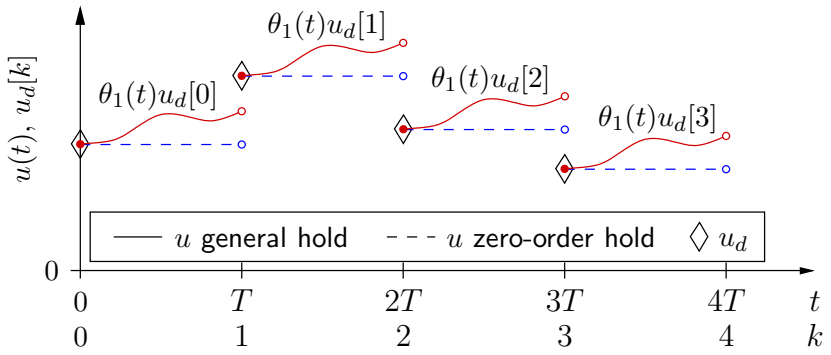

Figure 27 – Continuous-time system input $u(t)$ generated by a general hold device (solid lines) with hold function θ_1 and a zero-order hold (dashed lines) for the discrete-time control $u_d[k]$.

the vector $u(t) \in \mathbb{R}^p$. Figure 27 illustrates the operation of a general hold device for single-input systems. It shows that the continuous-time input $u(t)$ is obtained from concatenating the hold function repeatedly, where the discrete-time control instances $u_d[k]$ have the meaning of weights of the hold function. Apparently, the general hold device behaves in the special case $\theta_i \equiv 1$, $i = 1, 2, \ldots, p$, as a zero-order hold. It is the aim of the approach presented in the following to design the hold functions such that the residual dynamics is excited in a specific way, namely such that their contributions to the system output y become marginal at the sampling time instances. In consequence, these contributions to the sampled output sequence y_d vanish almost so that the residual dynamics hardly influence the compensator and thus the closed-loop behavior.

As in the previous chapter, the sampled-data control is designed for continuous-time state linear systems

$$\Sigma: \quad \dot{x}(t) = \mathcal{A}x(t) + \mathcal{B}u(t), \qquad t > 0, \quad x(0) = x_0 \in X \tag{5.6}$$

$$y(t) = \mathcal{C}x(t), \qquad t \geq 0 \tag{5.7}$$

that satisfy the Assumptions 2.1-2 and 4.1-2, and a standard sampling device is used for converting the continuous-time system output y to the corresponding discrete-time signal y_d by

$$y_d[k] = y(t_k), \qquad k \in \mathbb{N}_0. \tag{5.8}$$

In the following subsection state space representations of sampled-data systems with a general hold device are determined. A modal approximation is provided for the compensator design and a state space model of the corresponding residual dynamics

for the purpose of closed-loop analysis. Subsection 5.1.2 addresses the design of the hold functions. Finally, the resulting closed-loop dynamics and the impact of spillover are analyzed in Subsection 5.1.3.

5.1.1 Sampled-data systems with general hold devices

The sampled-data systems considered in Chapter 4 were derived from the continuous-time plant models under the assumption that a zero-order hold device was used. Now, a discrete-time model of the sampled-data system will be determined that involves a hold device of the general type, whose functionality is described by (5.3)–(5.4). Since a sampled-data approximation of Σ is used for the compensator design the modal approximation

$$\Sigma_n: \quad \dot{x}_n(t) = A_n x_n(t) + B_n u(t), \qquad t > 0, \quad x_n(0) = \mathcal{F}^{-1}\mathcal{P}x_0 \in \mathbb{C}^n \tag{5.9}$$

$$y_n(t) = C_n x_n(t), \qquad t \geq 0, \tag{5.10}$$

that was introduced in Section 2.2, is converted to a discrete-time model. For this purpose, the solution of the state equation (5.9) is represented as

$$x_n(t_{k+1}) = e^{A_n(t_{k+1}-t_k)} x_n(t_k) + \int_{t_k}^{t_{k+1}} e^{A_n(t_{k+1}-\tau)} B_n u(\tau) \mathrm{d}\tau, \qquad \forall\, k \in \mathbb{N}_0 \tag{5.11}$$

(see [86, Sec. 2.5]). Inserting (5.3) and $t_{k+1} - t_k = T$, and substituting τ by $\tau + t_k$ yields

$$x_n[k+1] = e^{A_n T} x_n[k] + \int_0^T e^{A_n(T-\tau)} B_n \Theta(\tau) \mathrm{d}\tau \; u_d[k], \qquad \forall\, k \in \mathbb{N}_0. \tag{5.12}$$

So, by introducing the matrices

$$A_{d,n} := e^{A_n T} \tag{5.13}$$

$$B_{d,n} := \int_0^T e^{A_n(T-\tau)} B_n \Theta(\tau) \mathrm{d}\tau \tag{5.14}$$

one arrives at the discrete-time model

$$\Sigma_n^d: \quad x_n[k+1] = A_{d,n} x_n[k] + B_{d,n} u_d[k], \quad k \in \mathbb{N}_0, \quad x_n[0] = \mathcal{F}^{-1}\mathcal{P}x_0 \in X \tag{5.15}$$

$$y_{d,n}[k] = C_n x_n[k], \qquad k \in \mathbb{N}_0 \tag{5.16}$$

of the sampled-data approximation, where (5.16) follows immediately from (5.10) and the sampling law (5.8). In the analog way also the sampled-data system that corresponds to the continuous-time residual dynamics

$$\Sigma_r: \quad \dot{x}_r(t) = \mathcal{A}_r x_r(t) + \mathcal{B}_r u(t), \qquad t > 0, \quad x_r(0) = (I - \mathcal{P})x_0 \in X_r \tag{5.17}$$

$$y_r(t) = \mathcal{C}_r x_r(t), \qquad t \geq 0 \tag{5.18}$$

(see (2.117)–(2.118)) can be determined, yielding the discrete-time model

$$\Sigma_r^d: \quad x_r[k+1] = \mathcal{A}_{d,r} x_r[k] + \mathcal{B}_{d,r} u[k], \quad k \in \mathbb{N}_0, \quad x_r[0] = (I - \mathcal{P})x_0 \in X_r \tag{5.19}$$

$$y_{d,r}[k] = \mathcal{C}_r x_r[k], \qquad k \in \mathbb{N}_0 \tag{5.20}$$

with

$$\mathcal{A}_{d,r} h = \mathcal{S}_r(T)h, \qquad \forall\, h \in X_r \tag{5.21}$$

$$\mathcal{B}_{d,r} v = \int_0^T \mathcal{S}_r(T-\tau)\mathcal{B}_r \Theta(\tau)\mathrm{d}\tau\, v, \qquad \forall\, v \in \mathbb{C}^p, \tag{5.22}$$

wherein $\mathcal{S}_r(t)$ is the C_0-semigroup generated by $\mathcal{A}_r$ (compare to Subsection 4.2.1). As in Subsection 4.2.1 one can show that the sampled-data approximation Σ_n^d and the sampled-data residual dynamics Σ_r^d are complementary w.r.t. the transfer behavior, *i.e.*,

$$y_{d,n}[k] + y_{d,r}[k] = y_d[k] = y(t_k), \qquad \forall\, k \in \mathbb{N}_0 \tag{5.23}$$

(see (4.66)). Comparison of the models Σ_n^d and Σ_r^d with those determined in Section 4.2.1 shows that the system operators and output operators are the same but $B_{d,n}$ and $\mathcal{B}_{d,r}$ are changed and depend on $\Theta(t) = \mathrm{diag}(\theta_i(t), \ldots, \theta_p(t))$ (see (5.14) and (5.22)). The freedom in the choice of the hold functions θ_i is used in the next subsection to keep the impact of the residual dynamics on the closed-loop behavior small so that hence the spillover is suppressed.

5.1.2 Design of the hold functions

It suggests itself to choose the sampling functions such that $\|\mathcal{B}_{d,r}\|$ becomes as small as possible because the excitation of the discrete-time residual dynamics Σ_r^d is then reduced due to (5.19) so that it impacts the closed-loop system less. This is described in a more quantitative way by Theorem 4.2-4 that can be applied here because only $B_{d,n}$ and $\mathcal{B}_{d,r}$ have been modified, compared to the closed-loop system considered

in Subsection 4.2.3. This theorem shows that the perturbation of the closed-loop spectrum becomes small if so is $\|\mathcal{B}_{d,r}\|$ because then also the perturbation operator $\tilde{\Delta}$ has a small norm due to (4.86). For this reason the relation (5.22) between Θ and $\mathcal{B}_{d,r}$ is analyzed with respect to the objective to minimize $\|\mathcal{B}_{d,r}\|$. However, besides $\mathcal{B}_{d,r}$ also $B_{d,n}$ depends on Θ according to (5.14). Therefore, the choice of the sampling functions is restricted by the requirement that $(A_{d,n}, B_{d,n})$ is controllable in order to ensure that an arbitrary eigenvalue assignment by the state feedback is possible. Thus, for determining the hold functions the following problem is considered.

Problem 5.1-1

Minimize $\|\mathcal{B}_{d,r}\|$ under the constraints

1. $\theta_i \in L_\infty(0,T),\ \ i = 1, 2, \ldots, p$
2. $(A_{d,n}, B_{d,n})$ is controllable. ◀

For analyzing this problem the solution $x_{r,i}$ of the state equation (5.17) is considered that results from the input

$$u(t) = e_i\theta_i(t) = \Theta(t)e_i \tag{5.24}$$

with e_i denoting the i-th canonical basis vector of $\mathbb{R}^p$, and vanishing initial state $x_{r,i}(0) = 0$. This solution, evaluated at the time instant $t = T$, is given by

$$x_{r,i}(T) = \int_0^T \mathcal{S}_r(T-\tau)\mathcal{B}_r\Theta(\tau)\mathrm{d}\tau\, e_i, \qquad i = 1, 2, \ldots, p \tag{5.25}$$

(see [46, Thm. 3.1.7, Def. 3.1.4]), where the subscript i stands for the index of the hold function that is used for the input. Comparing this with (5.22) yields

$$x_{r,i}(T) = \mathcal{B}_{d,r}e_i, \qquad i = 1, 2, \ldots, p. \tag{5.26}$$

This shows that the minimization of $\|\mathcal{B}_{d,r}e_i\|$ is achieved if and only if the state $x_{r,i}$ of the continuous-time residual dynamics Σ_r is steered by the input $u(t) = e_i\theta_i(t)$ from the initial state $x_{r,i}(0) = 0$ to a final state $x_{r,i}(T)$ with minimal norm, for which reason it is desirable to yield $x_{r,i}(T) = 0$. Thus, the design problem for the hold functions has the form of a feedforward control problem for the continuous-time residual dynamics Σ_r. It is a well-known fact that the considered state linear systems are not

exactly controllable[1] due to the finite number p of inputs (see [46, Thm. 4.1.5]). Thus, one cannot expect to achieve $x_{r,i}(T) = 0$ and thus a complete spillover suppression in general[2]. Instead of exact controllability the concept of *approximate controllability*[3] is adequate for state linear systems with a finite-rank input operator (see [46, Sec. 4.1]). The fact that a system is not exactly but approximately controllable has the meaning that no input signal $u \in L_2([0,\tau];\mathbb{R}^p)$ exists such that $\|x_{r,i}(T)\|_{X_r} = 0$ holds, but instead $\|x_{r,i}(T)\|_{X_r}$ can be made arbitrarily small. This leads to two important practical difficulties. Firstly, any algorithm for computing u that minimizes this norm cannot terminate since a smallest value of $\|x_{r,i}(T)\|_{X_r}$ does not exist. Secondly, the corresponding input signals u that result from reducing the norm further and further divergate, *i.e.*, the maximal absolute values $\max_{t\in[0,T]} |u(t)|$ exceed any bound. Of course, both shortcomings are unacceptable for an implementation which is why Problem 5.1-1 is not feasible in general. Another difficulty comes from the fact that the minimization problem is infinite-dimensional as long as $u \in L_\infty(0,T)$ is considered. In order to avoid these difficulties the hold functions are restricted in the following to the finite-dimensional set of *step functions* over the interval $[0,T]$ with N steps, *i.e.*,

$$\theta_i \in K_\Theta := \Big\{ \sum\nolimits_{j=1}^{N} \alpha_j \kappa_j \;\Big|\; \alpha_j \in \mathbb{R},\ j = 1,2,\ldots,N \Big\}, \qquad i = 1,2,\ldots,p \tag{5.27}$$

with $N \geq 1$ and the characteristic functions

$$\kappa_j(t) = \begin{cases} 1 & : \quad t \in \left[(j-1)\frac{T}{N},\, j\frac{T}{N}\right) \\ 0 & : \quad \text{else} \end{cases} \qquad \text{for} \qquad t \in [0,T],\ j = 1,2,\ldots,N. \tag{5.28}$$

Note, that the coefficients α_j in (5.27) have the meaning of the step heights (see Figure 28). These parameters will be determined in the following such that $\|\mathcal{B}_{d,r}\|$ becomes small. This task will be done by solving a finite-dimensional convex optimization problem. Once they have been computed, the α_j are constant at runtime of

[1] A system Σ on the state space X is said to be *exactly controllable* if for any $x_\tau \in X$ an input $u \in L_2([0,\tau];\mathbb{R}^p)$ with a finite $\tau > 0$ exists such that the corresponding system state x satisfies $x(0) = 0$ and $x(\tau) = x_\tau$ (see [46, Def. 4.1.3]).

[2] For some certain classes of parabolic and biharmonic systems it is possible to find feedforward controls u such that $x_{r,i}(T) = 0$, $i = 1,2,\ldots,p$, and thus $\mathcal{B}_{d,r} = 0$ so that the spillover is avoided entirely (see [19, 118]).

[3] A system Σ on the state space X is said to be *approximately controllable* if for any $x_\tau \in X$ and any $\varepsilon > 0$ an input $u \in L_2([0,\tau];\mathbb{R}^p)$ for finite $\tau > 0$ exists such that the corresponding system state x satisfies $x(0) = 0$ and $\|x(\tau) - x_\tau\|_X < \varepsilon$. If the system is approximately controllable but not exactly controllable, $\|u\|_{L_2}^2 = \int_0^\tau \|u(\tilde{\tau})\|_{\mathbb{R}^p}^2 \mathrm{d}\tilde{\tau}$ tends toward infinity for $\varepsilon \to 0$. Thus, $\varepsilon = 0$ cannot be yield due to the limitation to a finite amount of control energy (see [46, Def. 4.1.3]).

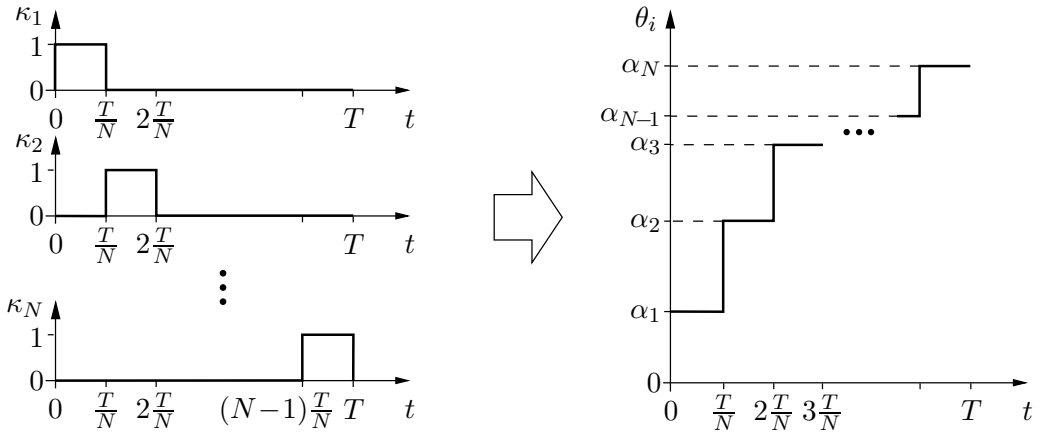

Figure 28 – Step function θ_i as characterized by the set K_Θ in (5.27), which consists of linear combinations of the characteristic functions κ_j (see (5.28)).

the control. The use of this type of hold functions has the advantage that the general hold device can be implemented in an easy way by using a zero-order hold that operates with the reduced sampling constant $T_H = T/N$. Thus, the resulting closed-loop system can be regarded as a *multirate* control system[4]. The second constraint of Problem 5.1-1, the controllability of $(A_{d,n}, B_{d,n})$, can be assured by choosing a matrix $B_{d,n}^{desired} \in \mathbb{C}^{n\times p}$ such that $(A_{d,n}, B_{d,n}^{desired})$ is controllable, and designing the hold functions such that $B_{d,n} = B_{d,n}^{desired}$. This approach, to assign the input matrix of the discrete-time approximation, offers the possibility to avoid not only uncontrollability but also that $(A_{d,n}, B_{d,n})$ is "almost not controllable", which means that the closed-loop performance could be achieved only by undesired large input amplitudes $|u(t)|$. This issue is explained more in detail in Remark 5.1-5. These considerations lead to the following relaxed problem formulation.

Problem 5.1-2

Choose the hold functions θ_i such that $\|\mathcal{B}_{d,r}\|$ is minimized under the constraints

1. $\theta_i \in K_\Theta,\ \ i = 1, 2, \ldots, p$
2. $B_{d,n} = B_{d,n}^{desired}$. ◀

[4] Multirate control systems are considered, *e.g.*, in [32, 119].

For solving this problem single-input systems are considered first so that (5.3) becomes

$$u(t) = \theta_1(t)u_d[k] = \sum_{j=1}^{N} \alpha_j \kappa_j(t) u_d[k], \qquad t \in [0, T), \tag{5.29}$$

wherein (5.27) is used. Then, the results are extended to the multi-input case subsequently. Let $x_{r,1}^j$, $j = 1, 2, \ldots, N$, denote the solutions of (5.17) that result when the residual dynamics Σ_r with initial state $x_{r,1}^j(0) = 0$ are excited by $u(t) = \kappa_j(t)$. Then, the solution $x_{r,1}$ of Σ_r with initial state $x_{r,1}(0) = 0$ that corresponds to the input u according to (5.29) can be written by superposition as

$$x_{r,1}(t) = \sum_{j=1}^{N} \alpha_j x_{r,1}^j(t) \tag{5.30}$$

due to the linearity of the system. By taking $\|\mathcal{B}_{d,r}\| = \|x_{r,1}(T)\|_{X_r}$ (see (5.26)) into account this shows that the minimization of $\|\mathcal{B}_{d,r}\|$ is equivalent to minimizing $\|\sum_{j=1}^{N} \alpha_j x_{r,1}^j(T)\|_{X_r}$ with $\alpha_j \in \mathbb{R}$. In order to reformulate the second constraint of Problem 5.1-2 one finds

$$x_{n,1}(T) = \int_0^T e^{A_n(T-\tau)} B_n \theta_1(\tau) \mathrm{d}\tau = B_{d,n} \tag{5.31}$$

in an analog way as (5.25)–(5.26). Similar to (5.30), the state trajectory $x_{n,1}$ of the approximation Σ_n that results from the input $u(t) = \sum_{j=1}^{N} \alpha_j \kappa_j(t)$ and the initial state $x_{n,1}(0) = 0$ can be represented as

$$x_{n,1}(t) = \sum_{j=1}^{N} \alpha_j x_{n,1}^j(t), \tag{5.32}$$

wherein the trajectories $x_{n,1}^j$, $j = 1, 2, \ldots, N$, correspond to the excitations of Σ_n by $u(t) = \kappa_j(t)$ and vanishing initial state $x_{n,1}^j(0) = 0$. Thus, Item 2 of Problem 5.1-2 becomes $\sum_{j=1}^{N} \alpha_j x_{n,1}^j(T) = B_{d,n}^{desired}$ due to (5.31)–(5.32). In summary, Problem 5.1-2 can be solved in the following way, leading to a design approach for the hold function θ_1 for systems with a single input.

Theorem 5.1-3

Assume that Σ has a single input $u(t) \in \mathbb{R}$ and let $x_{r,1}^j$ and $x_{n,1}^j$ denote the respective state trajectories of Σ_r and Σ_n for vanishing initial state and input $u(t) = \kappa_j(t)$. Suppose that $\alpha_1, \alpha_2, \ldots, \alpha_N$ solve the minimization problem

$$\min_{\alpha \in \mathbb{R}^N} \alpha^T M \alpha \tag{5.33}$$

with

$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}, \qquad M = \mathrm{Re} \begin{bmatrix} \langle x_{r,1}^1(T), x_{r,1}^1(T) \rangle_{X_r} & \cdots & \langle x_{r,1}^1(T), x_{r,1}^N(T) \rangle_{X_r} \\ \vdots & & \vdots \\ \langle x_{r,1}^N(T), x_{r,1}^1(T) \rangle_{X_r} & \cdots & \langle x_{r,1}^N(T), x_{r,1}^N(T) \rangle_{X_r} \end{bmatrix} \tag{5.34}$$

subject to

$$\begin{bmatrix} x_{n,1}^1(T) & x_{n,1}^2(T) & \cdots & x_{n,1}^N(T) \end{bmatrix} \alpha = B_{d,n}^{desired}. \tag{5.35}$$

Then, $\theta_1(t) = \sum_{j=1}^N \alpha_j \kappa_j(t)$ is the optimal hold function that solves Problem 5.1-2 for the single-input case.

The proof is given in Appendix A.12. Of course, the constraint (5.35) can be satisfied in general only if the states $x_{n,1}^j(T)$ span $\mathbb{C}^n$ for real coefficients α_j. This can be influenced by the number $N \geq n$ of steps of the hold functions. It is possible to assure that (5.35) has a (real) solution α by choosing the vector $B_{d,n}^{desired}$ equal to the input vector $B_{d,n}$ in (4.53) that results when a zero-order hold is used. Since a general hold with $\alpha_j = 1$, $j = 1, 2, \ldots, N$, is the same as a zero-order hold, this means that (5.35) has at least the solution $\alpha_j = 1$, and the optimization problem is thus feasible. If (5.35) has a solution, the optimization problem is convex because M is positive semidefinite and hence (5.33) a convex objective function, and (5.35) is an affine constraint (see [24]). Thus, every local minimum is a global one which can be computed, *e.g.*, by the solver `quadprog` of MathWorks' numerical computing software Matlab.

Remark 5.1-4 For determining $x_{n,1}^j(T)$ and $x_{r,1}^j(T)$, $j = 1, 2, \ldots, N$, the state equations (5.9) and (5.17) have to be solved for vanishing initial states and the corresponding input $u(t) = \kappa_j(t)$. Since the approximation Σ_n is described by a set of ODEs the computation of $x_{n,1}^j$ can be done by standard software tools. In contrast, the residual dynamics involve the same PDEs as the plant Σ so that a suitable solver has to be available for computing the state trajectories. Alternatively, it might be easier to consider a sufficiently accurate finite-dimensional approximation of the residual dynamics that is based on ODEs whose treatment is a standard issue. For reducing the computational costs it is useful to observe that it holds

$$x_{n,1}^j(t) = x_{n,1}^1\left(t - (j-1)T/N\right), \qquad j = 2, 3, \ldots, N \tag{5.36}$$

$$x_{r,1}^j(t) = x_{r,1}^1\left(t - (j-1)T/N\right), \qquad j = 2, 3, \ldots, N \tag{5.37}$$

due to the time-invariance of the considered systems, where $x^1_{n,1}(t) = x^1_{r,1}(t) = 0$ for $t < 0$. This shows that $x^j_{n,1}(T)$ and $x^j_{r,1}(T)$ have to be computed only for $j = 1$ while the remaining trajectories for $j = 2, 3, \ldots, N$ are yield from (5.36)–(5.37). ◀

Remark 5.1-5 As mentioned before, the freedom in the input vector $B^{desired}_{d,n}$ can be used to avoid that the approximation is hardly controllable, which leads to large amplitudes of the discrete-time control u_d. A particularly simple measure for quantifying the individual controllability of the single modes of the discrete-time approximation, that makes use of the diagonal shape of the dynamic matrix $A_{d,n}$ (see (2.138) and (5.13)), is given by

$$\mu_{c,i} := \|b_{d,n,i}\|_{\mathbb{C}^p}, \qquad i = 1, 2, \ldots, n, \tag{5.38}$$

where $b^T_{d,n,i}$ denotes the i-th row of $B_{d,n}$ (compare to [96]). This measure has the following meaning: Applying the state feedback $u_d[k] = -ke_i^T x_n[k]$ to the approximation enables to shift the eigenvalue $\lambda_{d,i}$ to $\tilde{\lambda}_{d,i}$, where the difference has the bound $|\lambda_{d,i} - \tilde{\lambda}_{d,i}| \leq \mu_{c,i}\|k\|_{\mathbb{C}^p}$. Thus, low numbers $\mu_{c,i}$ correspond to large feedback gains and thus large input amplitudes $|u_d[k]|$ for a specified eigenvalue shift. Therefore, it suggests itself to choose $B^{desired}_{d,n}$ such that the numbers $\mu_{c,i}$ are large and thus $|u_d[k]|$ is kept small. However, doing to leads to qualitatively large step heights α_j in view of (5.35) and (5.38), which has the consequence that the reduction of the amplitudes of u_d is compensated by a large gain of the general hold device. A more detailed analysis shows that the amplitudes of the finally resulting continuous-time input signal u depend on the quotients $\mu_{c,i}/\mu_{c,j}$, $i, j = 1, 2, \ldots, n$, which should not be excessively large in order to avoid large amplitudes.

The algorithm of Theorem 5.1-3 yields a minimization of $\|\mathcal{B}_{d,r}\|$. In addition, it may be desirable that the differences $\alpha_{j+1} - \alpha_j$, that describe the step height changes of the holds function θ_1, are not excessively large because otherwise the resulting input trajectories might be not realizable by the actor. These differences can be described by

$$\begin{bmatrix} \alpha_2 - \alpha_1 \\ \alpha_3 - \alpha_2 \\ \vdots \\ \alpha_N - \alpha_{N-1} \end{bmatrix} = D\alpha := \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \alpha. \tag{5.39}$$

Thus, for taking the step height differences into account in the design, (5.33) is replaced

by

$$\min_{\alpha\in\mathbb{R}^N} \alpha^T(M+\delta D^T D)\alpha, \tag{5.40}$$

where $\alpha^T D^T D\alpha$ describes the square of the norm of the difference vector, and $\delta > 0$ is a corresponding weight. ◀

Remark 5.1-6 If the system Σ has several inputs, *i.e.*, $u(t) \in \mathbb{R}^p$, $p > 1$, the approach of Theorem 5.1-3 has to be applied for each of the inputs. This means that the trajectories $x^1_{n,i}$ and $x^1_{r,i}$, $i = 1, 2, \ldots, p$, of Σ_n and Σ_r, respectively, have to be computed, that result from the input $u(t) = e_i\kappa_1(t)$ and vanishing initial states. By aid of Remark 5.1-4 the trajectories $x^j_{n,i}$ and $x^j_{r,i}$, $i = 1, 2, \ldots, p$, $j = 2, 3, \ldots, N$, that correspond to $u(t) = e_i\kappa_j(t)$ are obtained from

$$x^j_{n,i}(t) = x^1_{n,i}\left(t-(j-1)T/N\right), \qquad x^j_{r,i}(t) = x^1_{r,i}\left(t-(j-1)T/N\right). \tag{5.41}$$

Then, the minimization problems

$$\min_{\alpha_i\in\mathbb{R}^N} \alpha_i^T M_i\alpha_i, \qquad i = 1, 2, \ldots, p \tag{5.42}$$

with

$$\alpha_i = \begin{bmatrix}\alpha_{1,i}\\ \vdots \\ \alpha_{N,i}\end{bmatrix}, \qquad M_i = \mathrm{Re}\begin{bmatrix}\langle x^1_{r,i}(T), x^1_{r,i}(T)\rangle_{X_r} & \cdots & \langle x^1_{r,i}(T), x^N_{r,i}(T)\rangle_{X_r}\\ \vdots & & \vdots \\ \langle x^N_{r,i}(T), x^1_{r,i}(T)\rangle_{X_r} & \cdots & \langle x^N_{r,i}(T), x^N_{r,i}(T)\rangle_{X_r}\end{bmatrix} \tag{5.43}$$

subject to

$$\begin{bmatrix}x^1_{n,i}(T) & x^2_{n,i}(T) & \cdots & x^N_{n,i}(T)\end{bmatrix}\alpha_i = B^{desired}_{d,n}\, e_i, \qquad i = 1, 2, \ldots, p \tag{5.44}$$

have to be solved. Finally,

$$\theta_i(t) = \sum_{j=1}^{N} \alpha_{j,i}\kappa_j(t) \tag{5.45}$$

are the corresponding hold functions. However, instead of the operator norm $\|\mathcal{B}_{d,r}\| = \min_{\|v\|_{\mathbb{C}^p}\le 1}\|\mathcal{B}_{d,r}v\|_{X_r}$ the *Hilbert-Schmidt Norm* $\|\mathcal{B}_{d,r}\|_{HS} = \sqrt{\sum_{i=1}^p \|\mathcal{B}_{d,r}e_i\|^2_{X_r}}$ is minimized which comes from minimizing each of the norms $\|\mathcal{B}_{d,r}e_i\|_{X_r}$, $i = 1, 2, \ldots, p$, independently. Thus, Problem 5.1-2 is solved for multi-input systems in the sense that the minimization refers to the Hilbert-Schmidt norm. Note, that $\|\mathcal{B}_{d,r}\|$ and $\|\mathcal{B}_{d,r}\|_{HS}$ coincide in the single-input case, and one has $\|\mathcal{B}_{d,r}\| \le \|\mathcal{B}_{d,r}\|_{HS}$ in general. Thus, by minimizing the Hilbert-Schmidt norm also the operator norm becomes small. ◀

By minimizing the norm $\|\mathcal{B}_{d,r}\|_{HS}$ the excitation of the residual dynamics Σ_r and hence the control spillover become small, as has been argued in detail in Subsection 2.3.2 for the continuous-time control. In consequence, the perturbation of the closed-loop spectrum can be expected to be small. This will be analyzed in the following subsection.

5.1.3 Analysis of the closed-loop dynamics for observer-based control

As in Section 4.2 the sampled-data system Σ^d is controlled by an discrete-time observer-based compensator

$$\Sigma_c^d: \quad \hat{x}_n[k+1] = (A_{d,n} - LC_n)\hat{x}_n[k] + B_{d,n}u_d[k] + Ly_d[k], \qquad k \in \mathbb{N}_0 \tag{5.46}$$

$$\hat{x}_n[0] = \hat{x}_{n,0} \in \mathbb{C}^n \tag{5.47}$$

$$u_d[k] = -K\hat{x}_n[k], \qquad k \in \mathbb{N}_0 \tag{5.48}$$

(see Figure 26) that is designed on the basis of the discrete-time approximation Σ_n^d in (5.15)–(5.16), for which purpose the aspects of Subsection 4.2.2 for achieving a prescribed stability margin have to be taken into account. Consequently, the closed-loop spectrum is again described by (4.84)–(4.86). Furthermore, the results in Corollary 4.2-2 and Theorem 4.2-4 concerning the structure of the closed-loop spectrum and the spectrum perturbation can be applied. As discussed before, the use of the general hold device instead of the zero-order hold has the effect that the norm of $\mathcal{B}_{d,r}$ becomes small when the hold functions are designed as suggested by Theorem 5.1-3. Therefore, also the norm of the perturbation operator $\tilde{\Delta}$ has a small value in view of (4.86). An expression for $\mathcal{B}_{d,r}$ follows directly from (5.26), wherein the state $x_{r,i}$ of Σ_r, that corresponds to the input $u(t) = e_i\theta_i(t) = \sum_{j=1}^N \alpha_{j,i}\kappa_j(t)$, is given by

$$x_{r,i}(t) = \sum_{j=1}^{N} \alpha_{j,i} x_{r,i}^j(t), \qquad i = 1, 2, \ldots, p \tag{5.49}$$

in view of Remark 5.1-6 and the linearity of the system. Taking this in (5.26) into account gives

$$\mathcal{B}_{d,r}v = \sum_{j=1}^{N} \begin{bmatrix} \alpha_{j,1}x_{r,1}^j(T) & \alpha_{j,2}x_{r,2}^j(T) & \cdots & \alpha_{j,p}x_{r,p}^j(T) \end{bmatrix} v, \qquad \forall\, v \in \mathbb{C}^p. \tag{5.50}$$

It can be shown easily that it holds $\|\tilde{\Delta}\| \leq \sqrt{2}\|K\|\|\mathcal{B}_{d,r}\|$ for the perturbation operator $\tilde{\Delta}$ in (4.86). Consequently, Theorem 4.2-4 can be reformulated as follows.

Corollary 5.1-7

Let the Assumption 4.1-2 hold and assume that the observer-based compensator Σ_c^d in (5.46)–(5.48) is designed such that the eigenvalues of $A_{d,n} - B_{d,n}K$ and $A_{d,n} - LC_n$ are simple, mutually different, and not contained in $\sigma(\mathcal{A}_{d,r})$. Then,

$$d_d \leq \sqrt{2}\|\tilde{\mathcal{T}}_{cl}^{-1}\|\|\tilde{\mathcal{T}}_{cl}\|\|K\|\|\mathcal{B}_{d,r}\| \tag{5.51}$$

is an upper bound for the spectrum perturbation d_d (see (4.82)). Therein, $\mathcal{B}_{d,r}$ is given by (5.50), where the $x_{r,i}^j(T)$ result from the approach of Theorem 5.1-3 and Remark 5.1-6, and $\tilde{\mathcal{T}}_{cl}$ is a linear transformation such that $\tilde{\mathcal{T}}_{cl}^{-1}\tilde{\mathcal{A}}_{d,cl,0}\tilde{\mathcal{T}}_{cl}$ with $\tilde{\mathcal{A}}_{d,cl,0}$ from (4.86) is normal.

Thus, the influence of the residual dynamics on the closed-loop behavior can be suppressed by suitably choosing the hold functions such that the norms $\|\sum_{j=1}^N \alpha_{j,i}\, x_{r,i}^j(T)\|_{X_r}$, $i = 1, 2, \ldots, p$, are minimized. Note, that these norms can be made small in a trivial way by choosing $B_{d,n}^{desired}$ in (5.35) and (5.44) with a small norm. However, doing so leads to a large norm of K so that both effects cancel in the right hand-side of (5.51). Nevertheless, d_d can be reduced arbitrarily by increasing N, so that the spectrum perturbation becomes arbitrarily small, which is stated next.

Theorem 5.1-8

Let the assumptions of Corollary 5.1-7 be satisfied. Then, for every $\varepsilon > 0$ a number N of steps of the step functions in (5.27)– (5.28) exists such that the design approach of Theorem 5.1-3 and Remark 5.1-6 yields a spectrum perturbation d_d that satisfies $d_d < \varepsilon$.

The proof is given in Appendix A.13. In the following example the approach of this section is demonstrated.

Example 5.1-9 (Control of an Euler-Bernoulli beam with Kelvin-Voigt damping, continued)

In Example 4.2-6 a sampled-data control was designed for an Euler-Bernoulli beam with Kelvin-Voigt damping. The length of the beam is $\ell = 1$, the damping constant is $\delta = 0.005$, and the input and output distribution functions are given by (4.92).

As there, also here the most four dominant eigenvalues $\lambda_{\pm 1} = -0.487 \pm j9.86$ and $\lambda_{\pm 2} = -7.79 \pm j38.70$ of the system operator $\mathcal{A}$ in the continuous-time domain are shifted by assigning the eigenvalues $\lambda_{c,\pm 1} = -10 \pm j5$, $\lambda_{c,\pm 2} = -10 \pm j10$ of the controlled continuous-time approximation and $\lambda_{o,\pm 1} = -15 \pm j5$, $\lambda_{o,\pm 2} = -15 \pm j10$ of the continuous-time counterpart of the observer dynamics. In view of the sampling constant $T = 0.025$ this should yields a stability margin $\beta_d = 0.839$ for the discrete-time closed-loop system. The discrete-time approximation Σ_n^d determined in Example 4.2-6 (see (4.93)–(4.95)) is used. Thus, the discrete-time observer-based compensator Σ_c^d in (5.46)–(5.48) is the same as in Example 4.2-6. There, it was shown that the compensator in combination with the standard hold device and standard sampling did not yield the desired stability margin $\beta_d = 0.839$ for the discrete-time closed system operator due to the spillover impact. In order to reduce the spectrum perturbation a general hold device is designed in the following.

For applying the approach of Theorem 5.1-3 the state trajectory $x_{n,1}^1$ of the approximation Σ_n, that results from the input $u(t) = \kappa_1(t)$ (see (5.28)) and initial state $x_{n,1}^1(0) = 0$, is computed by numerical integration. The trajectories $x_{n,1}^j$, $j = 2, 3, \ldots, N$, are obtained subsequently by suitable time-shifts of $x_{n,1}^1$ (see Remark 5.1-4). For determining the trajectory $x_{r,1}^1$ of the residual dynamics Σ_r the state trajectory x_1^1 of the plant Σ is computed first, that results from the input $u(t) = \kappa_1(t)$ and initial state $x_1^1(0) = 0$. This is done by discretizing the spatial coordinate uniformly into 200 intervals and using the *method of lines* (see [95]). Then, $x_{r,1}^1$ is determined from

$$x_{r,1}^1(t) = (I - \mathcal{P})x_1^1(t) = x_1^1(t) - \mathcal{F}x_{n,1}^1(t), \qquad t \geq 0 \tag{5.52}$$

(see Section 2.2). Finally, $x_{r,1}^j$, $j = 2, 3, \ldots, N$, is obtained by time-shifting $x_{r,1}^1$ in view of Remark 5.1-4. For the choice of the vector $B_{d,n}^{desired}$ in (5.35) the vector

$$B_{d,n}^{desired} := \begin{bmatrix} -0.0029 + j0.0234 \\ -0.0286 - j0.2312 \\ -0.0059 + j0.0117 \\ -0.2332 - j0.4615 \end{bmatrix} \tag{5.53}$$

is used that coincides with the input vector $B_{d,n}$ in (4.94) that results when a zero-order hold is used. In this way it is assured that (5.35) can be fulfilled for $N = 1$ and hence for any $N \in \mathbb{N}$ at least by choosing $\alpha_{j,1} = 1$. Moreover, and $(A_{d,n}, B_{d,n}^{desired})$ is controllable and the quotients $\mu_{c,i}/\mu_{c,j}$, $i, j = 1, 2, \ldots, n$, are acceptable (see Remark 5.1-5). Since, finally, $(C_n, A_{d,n})$ is observable, the desired eigenvalues $\lambda_{d,c,i} = e^{\lambda_{c,i}T}$, $i = \pm 1, \pm 2$, can be assigned to $A_{d,n} - B_{d,n}K$ and $\lambda_{d,o,i} = e^{\lambda_{o,i}T}$ to $A_{d,n} - LC_n$. The hold functions

θ_1 with $N \in \{4, 10, 20\}$ steps that result from the minimization (5.42)–(5.44) are shown in Figure 29, and the corresponding closed-loop eigenvalue distributions are depicted in the Figures 30–32. The latter ones show that the perturbation of the desired eigenvalues becomes smaller when the number of steps N is increased, as can be expected in view of Theorem 5.1-8. This can be seen also from Table 6 in which the related spectrum perturbations d_d are given that have been determined on the basis of a modal approximation of the beam with order $n_{\text{high}} = 60$. The desired stability margin $\beta_d = 0.839$ is achieved for $N \in \{10, 20\}$ steps, as becomes apparent from the Figures 31–32.

In order to demonstrate the effect of taking the step height differences $\alpha_{i+1} - \alpha_i$ into account as suggested in Remark 5.1-5 the parameter δ in the modified objective function (5.40) is chosen $\delta \in \{0, 10^{-6}, 10^{-5}\}$, where $N = 20$ is considered. The resulting hold functions are depicted in Figure 33. As can be expected this figure shows that increasing δ leads to hold functions with reduced changes of the step heights, so that the required bandwidth of the actor is lowered. On the other hand, the spectrum perturbation d_d increases as δ is chosen larger. While $\delta = 0$ yields $d_d = 0.077$, one has $d_d = 0.104$ for $\delta = 10^{-6}$, and $d_d = 0.142$ for $\delta = 10^{-5}$. In all three cases the desired stability margin $\beta_d = 0.839$ is obtained. ◀

Table 6 – Spectrum perturbation d_d of the closed-loop system with a general hold device with $N = 1, 4, 10, 20$ steps that is yield by minimizing (5.33) subject to (5.35). The case $N = 1$ corresponds to the conventional zero-order hold. The Spectrum perturbation has been computated on the basis of a modal approximation of the beam with order $n_{\text{high}} = 60$.

N	1	4	10	20
d_d	0.329	0.220	0.136	0.077

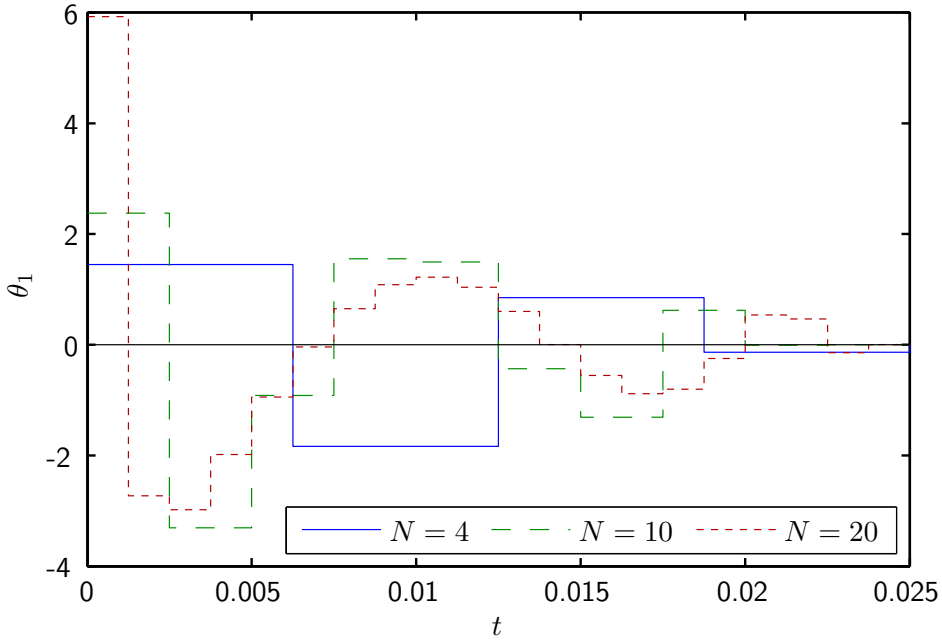

Figure 29 – Hold functions with $N = 4, 10, 20$ steps that minimize (5.33) subject to (5.35).

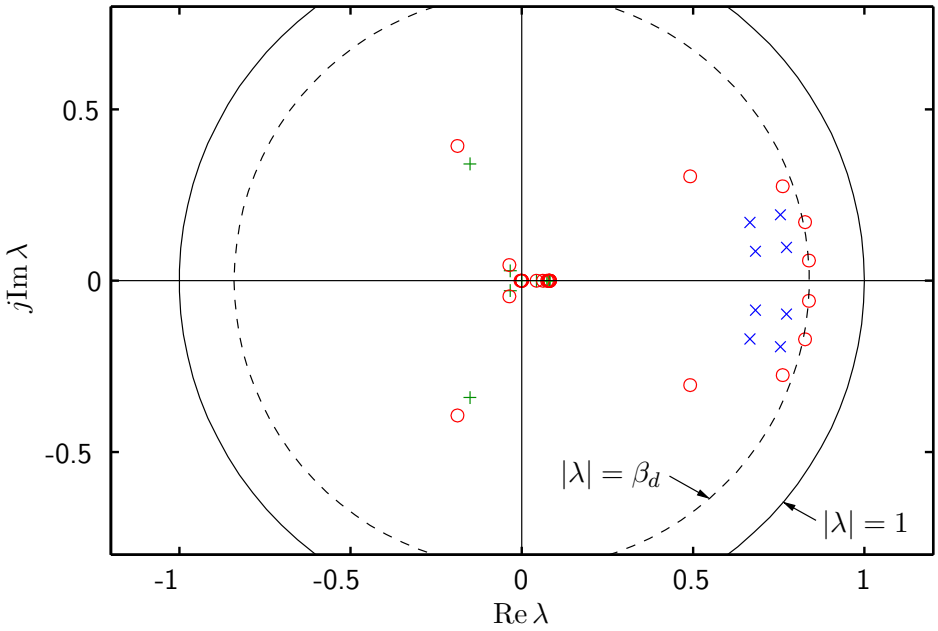

Figure 30 – Eigenvalues 'o' of the closed-loop system with a general hold device with $N = 4$ steps. For comparison the desired eigenvalues are shown, where 'x' marks $\lambda_{d,c,\pm1}$, $\lambda_{d,c,\pm2}$, $\lambda_{d,o,\pm1}$, $\lambda_{d,o,\pm2}$, and '+' describes $\sigma(\mathcal{A}_{d,r})$. The dashed circle defines the boundary of the region $\overline{\mathbb{C}}^i_{\beta_d}$ that corresponds to the stability margin $\beta_d = 0.839$.

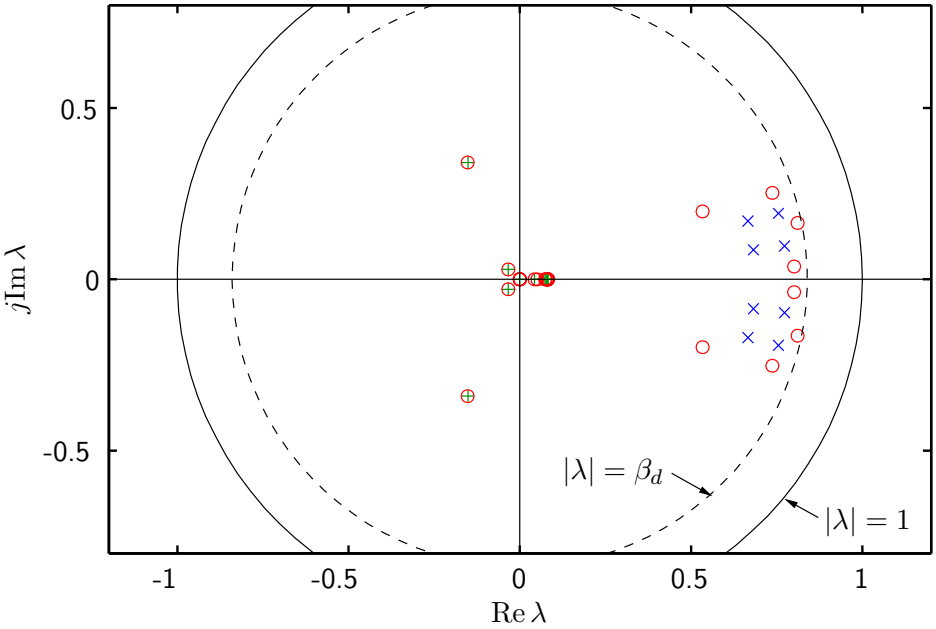

Figure 31 – Eigenvalues 'o' of the closed-loop system with a general hold device with $N = 10$ steps. For comparison the desired eigenvalues are shown, where 'x' marks $\lambda_{d,c,\pm1}$, $\lambda_{d,c,\pm2}$, $\lambda_{d,o,\pm1}$, $\lambda_{d,o,\pm2}$, and '+' describes $\sigma(\mathcal{A}_{d,r})$.

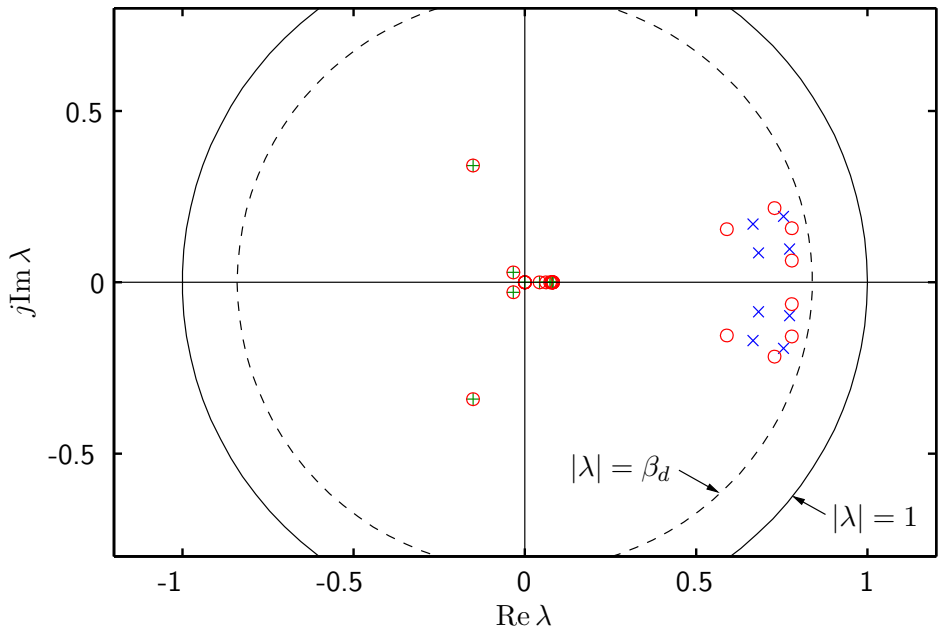

Figure 32 – Eigenvalues 'o' of the closed-loop system with a general hold device with $N = 20$ steps. For comparison the desired eigenvalues are shown, where 'x' marks $\lambda_{d,c,\pm1}$, $\lambda_{d,c,\pm2}$, $\lambda_{d,o,\pm1}$, $\lambda_{d,o,\pm2}$, and '+' describes $\sigma(\mathcal{A}_{d,r})$.

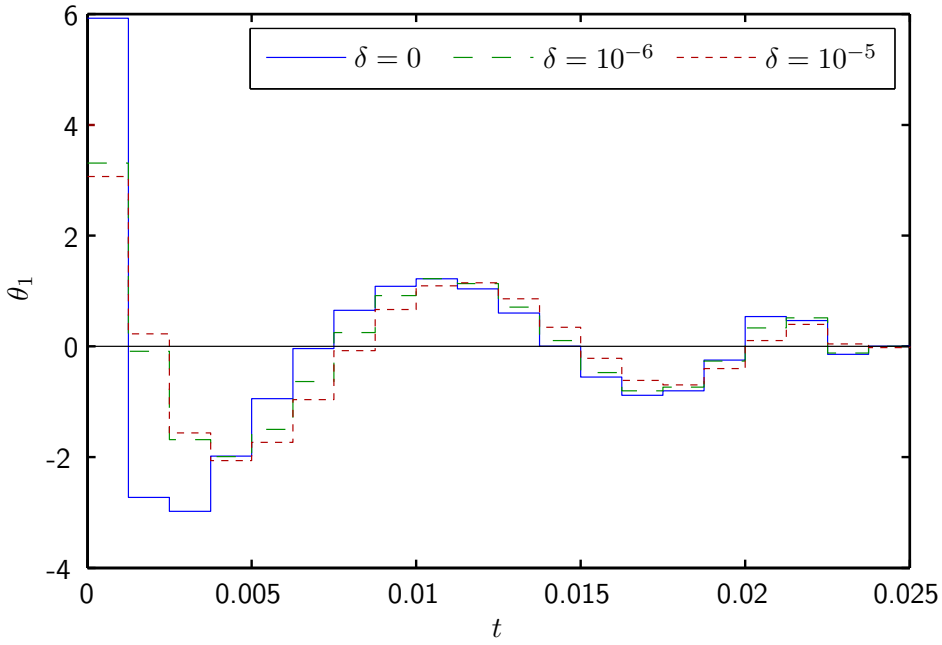

Figure 33 – Hold functions with $N = 20$ steps that minimize (5.40) with $\delta = 0, 10^{-6}, 10^{-5}$ subject to (5.35).

5.2 Control using general sampling

In this section a spillover suppression technique for sampled-data systems is considered whose basic concept is to use a general sampling device S_Π instead of standard sampling. As hold device the usual zero-order hold H is applied (see Figure 34). The general sampling device computes the instances $y_d[k]$ of the discrete-time output as the mean value of the system output $y(t)$ over the previous sampling interval, where, in doing so, the output is weighted by specifically designed *sampling functions* π_i. That means that its operation is described by

$$y_d[k] = \begin{cases} 0 & : \quad k = 0 \\ \begin{bmatrix} \int_{t_{k-1}}^{t_k} \pi_1(\tau - t_{k-1}) y_1(\tau) \mathrm{d}\tau \\ \vdots \\ \int_{t_{k-1}}^{t_k} \pi_m(\tau - t_{k-1}) y_m(\tau) \mathrm{d}\tau \end{bmatrix} & : \quad k \in \mathbb{N}, \end{cases} \tag{5.54}$$

wherein m is the number of outputs, *i.e.*, $y(t) \in \mathbb{R}^m$. The sampling functions

$$\pi_i : [0, T] \mapsto \mathbb{R}, \qquad i = 1, 2, \ldots, m \tag{5.55}$$

are considered at first as arbitrary functions. Later, however, they are restricted to a class of weighted linear combinations of Dirac delta functions since the integrals in (5.54) can then be replaced by sums. This has the consequence that the general sampling law can be implemented easily by means of standard sampling with a higher sampling rate. By introducing

$$\Pi(t) := \mathrm{diag}(\pi_1(t), \pi_2(t), \ldots, \pi_m(t)), \tag{5.56}$$

(5.54) can be simplified to

$$y_d[k] = \begin{cases} 0 & : \quad k = 0 \\ \int_{t_{k-1}}^{t_k} \Pi(\tau - t_{k-1}) y(\tau) \mathrm{d}\tau & : \quad k \in \mathbb{N}. \end{cases} \tag{5.57}$$

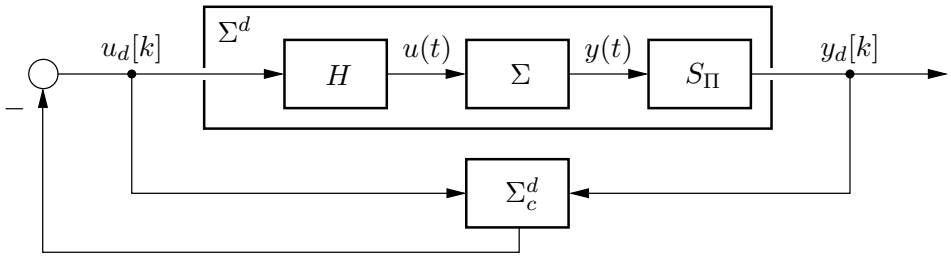

Figure 34 – Structure of a sampled-data control system that contains a zero-order hold device H and a general sampling device S_Π.

The sampling functions will be designed in the following such that the contributions of the residual dynamics to the discrete-time output y_d are suppressed as far as possible. In consequence, these dynamics hardly influence the compensator and thus the closed-loop behavior.

As before, continuous-time state linear systems

$$\Sigma: \quad \dot{x}(t) = \mathcal{A}x(t) + \mathcal{B}u(t), \qquad t > 0, \quad x(0) = x_0 \in X \tag{5.58}$$

$$y(t) = \mathcal{C}x(t), \qquad t \geq 0 \tag{5.59}$$

that satisfy the Assumptions 2.1-2 and 4.1-2 are the basis for the considerations in this section. In the first subsection state space representations of sampled-data systems with a zero-order hold, which is described by

$$u(t) = u_d[k], \qquad t \in [t_k, t_{k+1}),\ k \in \mathbb{N}_0, \tag{5.60}$$

and a general sampling device according to (5.57) are determined, where a modal approximation is determined for the compensator design and a state space model of the residual dynamics. In Subsection 5.2.2 an approach for the design of the sampling functions is presented. Finally, the resulting closed-loop dynamics and the impact of spillover are analyzed in Subsection 5.2.3.

5.2.1 Sampled-data systems with general sampling and observer-based control

It was shown in the previous section that the general hold device H_Θ impacts the state space models for the approximation and the residual dynamics such that only $B_{d,n}$ and $\mathcal{B}_{d,r}$ become changed. It is not surprising that the influence of the general sampling device is dual in the sense that the output matrix of the approximation and the output operator of the residual dynamics depend on the sampling functions π_i but the remaining operators are independent from π_i. Nevertheless, the situation with generalized sampling is more involved than the approach with a general hold device. To see this, it is important to observe from (5.57) and (5.59) that the discrete-time output y_d does not depend only on the current state $x(t_k)$ of the plant Σ but on the whole state trajectory $x(t)$, $t \in [t_{k-1}, t_k]$, over the last sampling interval. Clearly, this effect cannot be taken into account by simply modifying C_n and $\mathcal{C}_r$ since the output equations would then still depend only on $x(t_k)$. Instead, one has the following description of discrete-time approximation and the residual dynamics.

Proposition 5.2-1

Define the operator

$$\mathcal{B}_d(\tau)v := \int_0^\tau \mathcal{S}(\kappa)\mathrm{d}\kappa\, \mathcal{B}v, \qquad \tau \geq 0, \quad \forall\, v \in \mathbb{C}^p, \tag{5.61}$$

where $\mathcal{S}(t)$ is the C_0-semigroup generated by $\mathcal{A}$ that is given in (2.26), and consider the finite-dimensional system

$$x_n[k+1] = A_{d,n}x_n[k] + B_{d,n}u_d[k], \qquad k \in \mathbb{N}_0, \quad x_n[0] = \mathcal{F}^{-1}\mathcal{P}x_0 \in \mathbb{C}^n \tag{5.62}$$

$$y_{d,n}[k] = \begin{cases} 0 & : \quad k = 0 \\ \int_0^T \Pi(\tau)C_n e^{A_n\tau}\mathrm{d}\tau\, x_n[k-1] + \int_0^T \Pi(\tau)\mathcal{C}\mathcal{B}_d(\tau)\mathrm{d}\tau\, u_d[k-1] & : \quad k \in \mathbb{N} \end{cases} \tag{5.63}$$

and the infinite-dimensional system

$$x_r[k+1] = \mathcal{A}_{d,r}x_r[k] + \mathcal{B}_{d,r}u_d[k], \qquad k \in \mathbb{N}_0, \quad x_r[0] = (I-\mathcal{P})x_0 \in X_r \tag{5.64}$$

$$y_{d,r}[k] = \begin{cases} 0 & : \quad k = 0 \\ \int_0^T \Pi(\tau)\mathcal{C}_r\mathcal{S}_r(\tau)\mathrm{d}\tau\, x_r[k-1] & : \quad k \in \mathbb{N}. \end{cases} \tag{5.65}$$

Therein, $A_{d,n}$ and $B_{d,n}$ are the matrices of the discrete-time approximation Σ_n^d given in (4.52)–(4.53), and A_n and C_n are the matrices of the continuous-time approximation Σ_n (see (2.138) and (2.140)). Analog, $\mathcal{A}_{d,r}$ and $\mathcal{B}_{d,r}$ are the operators of the discrete-time residual dynamics Σ_r^d given in (4.60)–(4.61), $\mathcal{C}_r$ is the output operator of the continuous-time residual dynamics Σ_r (see (2.142)), and $\mathcal{S}_r(t)$ is the C_0-semigroup generated by $\mathcal{A}_r$ that is given in (4.62).

Then, the output y_d of the sampled-data system with input u_d, that consists of the continuous-time plant Σ, a zero-order hold device H, and a general sampling device S_Π, satisfies $y_d[k] = y_{d,n}[k] + y_{d,r}[k]$, $\forall\, k \in \mathbb{N}_0$.

The proof can be found in Appendix A.14. This shows that the systems (5.62)–(5.63) and (5.64)–(5.65) are complementary w.r.t. their outputs, *i.e.*, $y_{d,n}[k] + y_{d,r}[k] = y_d[k]$, $\forall\, k \in \mathbb{N}_0$, holds. Therefore, the first system can be regarded as an approximation for the sampled-data system and the second system describes the corresponding residual dynamics.

Remark 5.2-2 The integral in (5.61) can be evaluated by help of the representation

$$\int_0^\tau \mathcal{S}(\kappa)\mathrm{d}\kappa\, h = \sum_{i=1}^\infty \int_0^\tau e^{\lambda_i\kappa}\mathrm{d}\kappa\, \langle h, \psi_i\rangle_X \phi_i, \qquad \forall\, h \in X \tag{5.66}$$

(see (2.26)) and using

$$\int_0^\tau e^{\lambda_i \kappa} \mathrm{d}\kappa = \begin{cases} \frac{1}{\lambda_i}(e^{\lambda_i \tau} - 1) & : \quad \lambda_i \neq 0 \\ \tau & : \quad \lambda_i = 0. \end{cases} \tag{5.67}$$

In a similar way the integrals in (5.63) and (5.65) can be evaluated by considering the integrals for the single eigenmodes. ◀

Remark 5.2-3 Observe, that standard sampling is obtained from the general sampling in the special case $\pi_i(t) = \delta_T(t)$, $i = 1, 2, \ldots, m$, where δ_T denotes the Dirac delta function[5] centered at T, since insertion of these sampling functions into (5.54) yields $y_d[k] = y(t_k)$. One can show that in this case

$$y_d[k] = y_{d,n}[k] + y_{d,r}[k] = \mathcal{C}x[k], \qquad k \in \mathbb{N} \tag{5.68}$$

holds, which, as expected, coincides with the output equation found in (4.14). ◀

Note, that the state equations (5.62) and (5.64) are the same as in (4.50) and (4.58). This is not surprising since the latter ones were determined for the use of a zero-order hold that is considered also here. In consequence, $x_n[k]$ coincides with the state $x_n(t_k)$ of the continuous-time modal approximation Σ_n. However, the equations (5.62)–(5.63) do not constitute a state space model of the usual form since the output $y_{d,n}$ depends on $x_n[k-1]$ and $u_d[k-1]$, which means that the input-output relation contains dynamics that are not covered by the state equation. Hence, this model is not suitable as a basis for the compensator design. For that reason, the additional states $x_{n,-1}[k] := x_n[k-1]$ and $x_{u,-1}[k] := u_d[k-1]$ are introduced and the extended state

$$\bar{x}_n[k] := \begin{bmatrix} x_n[k] \\ x_{n,-1}[k] \\ x_{u,-1}[k] \end{bmatrix} = \begin{bmatrix} x_n[k] \\ x_n[k-1] \\ u_d[k-1] \end{bmatrix} \in \bar{X}_n := \mathbb{C}^n \oplus \mathbb{C}^n \oplus \mathbb{C}^p \tag{5.69}$$

is considered in the following. This leads to the discrete-time approximation

$$\bar{\Sigma}_n^d: \quad \bar{x}_n[k+1] = \bar{A}_{d,n}\bar{x}_n[k] + \bar{B}_{d,n}u_d[k], \qquad k \in \mathbb{N}_0 \tag{5.70}$$

$$\bar{x}_n[0] = \begin{bmatrix} \mathcal{F}^{-1}\mathcal{P}x_0 & 0 & 0 \end{bmatrix}^T \in \bar{X}_n \tag{5.71}$$

$$y_{d,n}[k] = \bar{C}_{d,n}\bar{x}_n[k], \qquad k \in \mathbb{N}_0 \tag{5.72}$$

[5] That means that δ_{t_0}, $t_0 \in [0,T]$, has in this context the defining properties $\delta_{t_0}(t) = 0$ for all $t \in [0,T] \setminus t_0$ and $\int_0^T \delta_{t_0}(\tau)\eta(\tau)\mathrm{d}\tau = \eta(t_0)$ for any continuous function $\eta : [0,T] \mapsto \mathbb{R}$.

with

$$\bar{A}_{d,n} = \begin{bmatrix} A_{d,n} & 0 & 0 \\ I & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{5.73}$$

$$\bar{B}_{d,n} = \begin{bmatrix} B_{d,n} \\ 0 \\ I \end{bmatrix} \tag{5.74}$$

$$\bar{C}_{d,n} = \begin{bmatrix} 0 & \int_0^T \Pi(\tau) C_n e^{A_n \tau} \mathrm{d}\tau & \int_0^T \Pi(\tau) \mathcal{C}\mathcal{B}_d(\tau) \mathrm{d}\tau \end{bmatrix} \tag{5.75}$$

by help of (5.62)–(5.63). Similar, the extension

$$\bar{x}_r[k] := \begin{bmatrix} x_r[k] \\ x_{r,-1}[k] \end{bmatrix} := \begin{bmatrix} x_r[k] \\ x_r[k-1] \end{bmatrix} \in \bar{X}_r := X_r \oplus X_r \tag{5.76}$$

yields the discrete-time residual dynamics

$$\bar{\Sigma}_r^d : \quad \bar{x}_r[k+1] = \bar{\mathcal{A}}_{d,r} \bar{x}_r[k] + \bar{\mathcal{B}}_{d,r} u_d[k], \qquad k \in \mathbb{N}_0 \tag{5.77}$$

$$\bar{x}_r[0] = \begin{bmatrix} (I - \mathcal{P}) x_0 \\ 0 \end{bmatrix} \in X_r \oplus X_r \tag{5.78}$$

$$y_{d,r}[k] = \bar{\mathcal{C}}_{d,r} \bar{x}_r[k], \qquad k \in \mathbb{N}_0 \tag{5.79}$$

with

$$\bar{\mathcal{A}}_{d,r} = \begin{bmatrix} \mathcal{A}_{d,r} & 0 \\ I & 0 \end{bmatrix} \tag{5.80}$$

$$\bar{\mathcal{B}}_{d,r} = \begin{bmatrix} \mathcal{B}_{d,r} \\ 0 \end{bmatrix} \tag{5.81}$$

$$\bar{\mathcal{C}}_{d,r} = \begin{bmatrix} 0 & \int_0^T \Pi(\tau) \mathcal{C}_r \mathcal{S}_r(\tau) \mathrm{d}\tau \end{bmatrix} \tag{5.82}$$

under use of (5.64)–(5.65). Since $\bar{\mathcal{A}}_{d,r}$ is bounded as $\mathcal{A}_{d,r}$ is, the difference equation (5.77) has a unique solution (see Subsection 4.1.2). Note, that $\sigma(\bar{\mathcal{A}}_{d,r})$, compared to $\sigma(\mathcal{A}_{d,r})$, has an additional eigenvalue $\lambda_{d,0} = 0$ whose multiplicity is infinite, *i.e.*,

$$\sigma_p(\bar{\mathcal{A}}_{d,r}) = \sigma_p(\mathcal{A}_{d,r}) \cup \{0\}. \tag{5.83}$$

Apparently, this additional spectral point is irrelevant for the stabilizability of the closed-loop system (see Assumption 4.2-1) since it is contained in $\overline{\mathbb{C}}_{\beta_d}^i$ for any $\beta_d \geq 0$. It is straightforward to verify that $(\bar{A}_{d,n}, \bar{B}_{d,n})$ is controllable if and only if $(A_{d,n}, B_{d,n})$ is controllable. The observability of $(\bar{C}_{d,n}, \bar{A}_{d,n})$, in contrast, depends on Π and thus on the sampling functions. This issue will be addressed later.

The discrete-time approximation $\bar{\Sigma}_n^d$ is used in the following as the basis for the design of an observer-based compensator $\bar{\Sigma}_c^d$ that, analog to (5.46)–(5.48), is described by

$$\bar{\Sigma}_c^d: \quad \hat{\bar{x}}_n[k+1] = (\bar{A}_{d,n} - L\bar{C}_{d,n})\hat{\bar{x}}_n[k] + \bar{B}_{d,n}u_d[k] + Ly_d[k], \qquad k \in \mathbb{N}_0 \tag{5.84}$$

$$\hat{\bar{x}}_n[0] = \hat{\bar{x}}_{n,0} \in \bar{X}_n \tag{5.85}$$

$$u_d[k] = -K\hat{\bar{x}}_n[k], \qquad k \in \mathbb{N}_0. \tag{5.86}$$

Note, that the controller order is $n_c = 2n + p$ instead of the order $n_c = n$ of the previously discussed compensators due to the extension of the state space. The considerations concerning the assignment of a desired stability margin, that have been discussed in Subsection 4.2.2, are relevant also here in analog form.

It becomes apparent from (5.65) that $y_{d,r}$ depends on the sampling functions π_i in view of $\Pi(t) = \mathrm{diag}(\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$. The freedom in the choice of them will be used subsequently with the aim to suppress $y_{d,r}$ since this contribution to the discrete-time output y_d excites the observer dynamics and thus leads to spillover (see Subsection 2.3.2 for the detailed discussion in continuous-time).

5.2.2 Design of the sampling functions

It is clear that the spillover is avoided entirely if the operator

$$\mathcal{C}_{d,r}h := \int_0^T \Pi(\tau)\mathcal{C}_r\mathcal{S}_r(\tau)\mathrm{d}\tau\, h, \qquad \forall\, h \in X_r \tag{5.87}$$

in (5.82) vanishes so that $y_{d,r}[k] = \bar{\mathcal{C}}_{d,r}\bar{x}_r[k] = 0$ holds, because then the residual dynamics $\bar{\Sigma}_r^d$ have no influence on the closed-loop system. Therefore, it is the aim in this subsection to design $\Pi(t)$ such that $\|\mathcal{C}_{d,r}\|$ becomes small. To this end, the adjoint of $\mathcal{C}_{d,r}$ is considered that reads

$$\mathcal{C}_{d,r}^* v = \int_0^T \mathcal{S}_r^*(\tau)\mathcal{C}_r^*\Pi^*(\tau)\mathrm{d}\tau\, v, \qquad \forall\, v \in \mathbb{C}^m. \tag{5.88}$$

By introducing

$$\tilde{\Pi}(\tau) := \Pi^*(T-\tau) = \mathrm{diag}(\pi_1(T-\tau), \pi_2(T-\tau), \ldots, \pi_m(T-\tau)) \tag{5.89}$$

(5.88) becomes

$$\mathcal{C}_{d,r}^* v = \int_0^T \mathcal{S}_r^*(\tau)\mathcal{C}_r^*\tilde{\Pi}(T-\tau)\mathrm{d}\tau\, v = \int_0^T \mathcal{S}_r^*(T-\tau)\mathcal{C}_r^*\tilde{\Pi}(\tau)\mathrm{d}\tau\, v, \qquad \forall\, v \in \mathbb{C}^m, \tag{5.90}$$

wherein τ is substituted by $T-\tau$ for the right term. Observe that this is the *dual equation*[6] of

$$\mathcal{B}_{d,r}v = \int_0^T \mathcal{S}_r(T-\tau)\mathcal{B}_r\Theta(\tau)\mathrm{d}\tau\, v, \qquad \forall\, v \in \mathbb{C}^p \tag{5.91}$$

that was found in (5.22), wherein $\tilde{\Pi}$ in (5.90) plays the role of $\Theta(t)$ in (5.91). In Subsection 5.1.2 an approach was presented for the design of $\Theta(t)$ such that $\|\mathcal{B}_{d,r}\|$ becomes small. Since $\|\mathcal{C}_{d,r}^*\|$ shall be minimized here, it suggests itself to apply again this method because (5.90) and (5.91) have the same form. To this end, the *dual residual dynamics*

$$\dot{\tilde{x}}_r(t) = \mathcal{A}_r^*\tilde{x}_r(t) + \mathcal{C}_r^*u(t), \qquad t>0, \quad \tilde{x}_r(0)=0 \tag{5.92}$$

is considered for which the input signals $u(t) = \tilde{\Pi}(t)e_i = e_i\pi_i(T-t)$, $i=1,2,\ldots,m$, yield the state trajectories

$$\tilde{x}_{r,i}(t) = \int_0^t \mathcal{S}_r^*(t-\tau)\mathcal{C}_r^*\tilde{\Pi}(\tau)e_i\mathrm{d}\tau, \qquad i=1,2,\ldots,m \tag{5.93}$$

for the initial conditions $\tilde{x}_{r,i}(0)=0$ (see [46, Thm. 3.1.7, Def. 3.1.4]). Thus,

$$\mathcal{C}_{d,r}^*e_i = \int_0^T \mathcal{S}_r^*(T-\tau)\mathcal{C}_r^*\tilde{\Pi}(\tau)e_i\mathrm{d}\tau, \qquad i=1,2,\ldots,m, \tag{5.94}$$

which follows from (5.90), coincides with the state $\tilde{x}_{r,i}(T)$ of the dual residual dynamics when excited by $u(t)=e_i\pi_i(T-t)$, *i.e.*,

$$\mathcal{C}_{d,r}^*e_i = \tilde{x}_{r,i}(T), \qquad i=1,2,\ldots,m. \tag{5.95}$$

The objective to minimize $\|\mathcal{C}_{d,r}^*\|$ can be regarded therefore as the feedforward control problem to find π_i such that $\|\tilde{x}_{r,i}(T)\|_{X_r}$ is as small as possible. To solve this problem, impulse functions are considered for the sampling functions, *i.e.*,

$$\pi_i \in K_\Pi := \Big\{\sum\nolimits_{j=1}^N \alpha_j\kappa_j \;\Big|\; \alpha_j\in\mathbb{R},\ j=1,2,\ldots,N\Big\}, \qquad i=1,2,\ldots,m \tag{5.96}$$

with $N\geq 1$ and

$$\kappa_j(t) = \delta_{(j-1)\frac{T}{N}}(t), \qquad t\in[0,T],\ j=1,2,\ldots,N, \tag{5.97}$$

[6] In the context of linear system theory an equation E_1 is called the *dual equation* of another equation E_2 if E_1 is obtained from E_2 by substituting a system operator $\tilde{\mathcal{A}}$ and the corresponding C_0-semigroup $\tilde{\mathcal{S}}(t)$, an input operator $\tilde{\mathcal{B}}$, and an output operator $\tilde{\mathcal{C}}$ in E_2 by $\tilde{\mathcal{A}}^*$, $\tilde{\mathcal{S}}^*(t)$, $\tilde{\mathcal{C}}^*$, and $\tilde{\mathcal{B}}^*$, respectively.

where δ_{t_0} denotes the Dirac delta function centered at t_0. A general sampling device with such sampling functions can be implemented comparatively simple by using standard sampling with a higher sampling rate. To see this, note that insertion of

$$\pi_i(t) = \sum_{j=1}^{N} \alpha_{j,i} \delta_{(j-1)\frac{T}{N}}(t) \tag{5.98}$$

into the sampling law (5.54) yields[7]

$$y_d[k] = \begin{cases} 0 & : \quad k = 0 \\ \begin{bmatrix} \sum_{j=1}^{N} \alpha_{j,1}\, y_1\big((j-1)T/N + (k-1)T\big) \\ \vdots \\ \sum_{j=1}^{N} \alpha_{j,m}\, y_m\big((j-1)T/N + (k-1)T\big) \end{bmatrix} & : \quad k \in \mathbb{N}. \end{cases} \tag{5.99}$$

Thus, the integrations in the general sampling law are replaced by weighted sums gained by standard sampling with the smaller sampling constant $T_S = T/N$, as it is known from the concept of multirate control (see, *e.g.*, [32, 119]). A further advantage of restricting the sampling functions to a finite-dimensional function space is that the minimization of $\|\mathcal{C}_{d,r}^*\|$ becomes a finite-dimensional problem.

Besides the minimization of the norm of $\mathcal{C}_{d,r}^*$ it has to be avoided that the discrete-time approximation $\bar{\Sigma}_n^d$ becomes unobservable since its output matrix $\bar{C}_{d,n}$ depends on $\Pi(t)$ in view of (5.75). More precisely, it suffices that the subsystem of $\bar{\Sigma}_n^d$ that corresponds to the states x_n and $x_{n,-1}$ is observable because the subsystem that corresponds to $x_{u,-1}$ has only eigenvalues at $\lambda_d = 0$ so that these need not to be changed by the compensator. For ensuring this, the sampling functions shall fulfill the additional requirement that the matrix

$$C_{d,n} := \int_0^T \Pi(\tau) C_n e^{A_n \tau} \mathrm{d}\tau \tag{5.100}$$

in (5.75) coincides with a specified output matrix $C_{d,n}^{desired}$. Thus, the following problem is considered.

Problem 5.2-4

Choose the sampling functions π_i such that $\|\mathcal{C}_{d,r}^*\|_{HS}$ is minimized under the constraints

1. $\pi_i \in K_\Pi, \;\; i = 1, 2, \ldots, m$

[7] It is used that y is continuous (see [46, Lem. 3.1.5]).

2. $C_{d,n} = C_{d,n}^{desired}$.

◀

This problem can be solved by an approach that is analog to that of Theorem 5.1-3. Note, that it is claimed to minimize the Hilbert-Schmidt norm $\|\mathcal{C}_{d,r}^*\|_{HS} = \sqrt{\sum_{i=1}^m \|\mathcal{C}_{d,r}^* e_i\|_{X_r}^2}$ instead of the operator norm $\|\mathcal{C}_{d,r}^*\| = \min_{\|v\|_{\mathbb{C}^m} \leq 1} \|\mathcal{C}_{d,r}^* v\|_{X_r}$ since Theorem 5.1-3 minimizes the Hilbert-Schmidt norm in the multi-input case (see Remark 5.1-6). To solve Problem 5.2-4, the *dual approximation*

$$\dot{\tilde{x}}_n(t) = A_n^* \tilde{x}_n(t) + C_n^* u(t), \qquad t > 0, \quad \tilde{x}_n(0) = 0 \tag{5.101}$$

is introduced. Then, the following design method for the sampling functions can be shown in an analog way as Theorem 5.1-3 when Remark 5.1-6 is taken into account. The basic difference is that the considerations refer now to the dual approximation (5.101) and the dual residual dynamics (5.92) instead of Σ_n and Σ_r.

Theorem 5.2-5

Let $\tilde{x}_{r,i}^j$ and $\tilde{x}_{n,i}^j$ denote the state trajectories that solve (5.92) and (5.101), respectively, for $u(t) = e_i \kappa_j(T-t)$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, N$. Suppose that the vectors $\alpha_i = \begin{bmatrix} \alpha_{1,i} & \alpha_{2,i} & \cdots & \alpha_{N,i} \end{bmatrix}^T$, $i = 1, 2 \ldots, m$, solve the minimization problems

$$\min_{\alpha_i \in \mathbb{R}^N} \alpha_i^T M_i \alpha_i, \qquad i = 1, 2, \ldots, m \tag{5.102}$$

with

$$M_i = \mathrm{Re} \begin{bmatrix} \langle \tilde{x}_{r,i}^1(T), \tilde{x}_{r,i}^1(T) \rangle_{X_r} & \cdots & \langle \tilde{x}_{r,i}^1(T), \tilde{x}_{r,i}^N(T) \rangle_{X_r} \\ \vdots & & \vdots \\ \langle \tilde{x}_{r,i}^N(T), \tilde{x}_{r,i}^1(T) \rangle_{X_r} & \cdots & \langle \tilde{x}_{r,i}^N(T), \tilde{x}_{r,i}^N(T) \rangle_{X_r} \end{bmatrix} \tag{5.103}$$

subject to

$$\begin{bmatrix} \tilde{x}_{n,i}^1(T) & \tilde{x}_{n,i}^2(T) & \cdots & \tilde{x}_{n,i}^N(T) \end{bmatrix} \alpha_i = (C_{d,n}^{desired})^* e_i, \qquad i = 1, 2, \ldots, m. \tag{5.104}$$

Then, $\pi_i(t) = \sum_{j=1}^N \alpha_{j,i} \kappa_j(t)$, $i = 1, 2, \ldots, m$, are the optimal sampling functions that solve Problem 5.2-4.

Similar to Theorem 5.1-3, the constraints (5.104) can be satisfied in general only if the states $\tilde{x}_{n,i}^j(T)$ span $\mathbb{C}^n$ for real coefficients $\alpha_{i,j}$. This can be influenced by the number

$N \geq n$. In Example 5.2-10 an approach is shown how it can be assured that (5.104) has a (real) solution α_i so that the optimization problems are feasible. If they are feasible, they are also convex because the matrices M_i are positive semidefinite which is why (5.102) describes convex objective functions, and the constraints in (5.104) are affine. Thus, every local minimum is a global one (see [24]) which can be computed, *e.g.*, by MATLAB's convex programming solver `quadprog`.

Remark 5.2-6 For determining $\tilde{x}^j_{n,i}(T)$ and $\tilde{x}^j_{r,i}(T)$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, N$, the state equations (5.101) and (5.92) of the dual approximation and the dual residual dynamics have to be solved for vanishing initial states and the corresponding input signals $u(t) = e_i\kappa_j(T-t) = e_i\delta_{(j-1)\frac{T}{N}}(T-t)$. Thus, $\tilde{x}^j_{n,i}$ and $\tilde{x}^j_{r,i}$ are time-shifted impulse responses of the dual approximation (5.101) and the dual residual dynamics[8] (5.92). Under use of

$$\tilde{x}^j_{n,i}(t) = e^{A_n^* t}\tilde{x}^j_{n,i}(0) + \int_0^t e^{A_n(t-\tau)}C_n^* u(\tau)\mathrm{d}\tau \tag{5.105}$$

$$\tilde{x}^j_{r,i}(t) = \mathcal{S}_r^*(t)\tilde{x}^j_{r,i}(0) + \int_0^t \mathcal{S}_r^*(t-\tau)\mathcal{C}_r^* u(\tau)\mathrm{d}\tau, \tag{5.106}$$

wherein $\mathcal{S}_r^*(t)$ is the C_0-semigroup generated by $\mathcal{A}_r^*$ (see [86, Sec. 2.5] and [46, Thm. 3.1.7]), one obtains for $\tilde{x}^j_{n,i}(0) = \tilde{x}^j_{r,i}(0) = 0$ and $u(t) = e_i\delta_{(j-1)\frac{T}{N}}(T-t)$

$$\tilde{x}^j_{n,i}(t) = e^{A_n^*(t-T+(j-1)\frac{T}{N})}C_n^* e_i \tag{5.107}$$

$$\tilde{x}^j_{r,i}(t) = \mathcal{S}_r^*(t-T+(j-1)\tfrac{T}{N})\mathcal{C}_r^* e_i. \tag{5.108}$$

Comparison with (5.105)–(5.106) shows that especially for $j = N+1$ the trajectories $x^{N+1}_{n,i}$ and $x^{N+1}_{r,i}$ can equivalently be obtained from

$$\dot{\tilde{x}}^{N+1}_{n,i}(t) = A_n^*\tilde{x}^{N+1}_{n,i}(t), \qquad t > 0, \quad \tilde{x}^{N+1}_{n,i}(0) = C_n^* e_i \tag{5.109}$$

$$\dot{\tilde{x}}^{N+1}_{r,i}(t) = \mathcal{A}_r^*\tilde{x}^{N+1}_{r,i}(t), \qquad t > 0, \quad \tilde{x}^{N+1}_{r,i}(0) = \mathcal{C}_r^* e_i. \tag{5.110}$$

Thus, instead of considering the impulse excitation the homogeneous (abstract) initial value problems (5.109)–(5.110) with appropriate initial states can be considered. The system (5.109) of ODEs can be solved by standard software tools, and the solution of

[8] Note, that despite the impulse excitation the resulting state trajectories are well-defined since a mild solution exists for $u \in L_\rho([0,\tau],\mathbb{R}^p)$ for a $\rho \in \mathbb{N}$, $\tau > 0$ (see [46, Lem. 3.1.5]). This is satisfied for $\rho = 1$.

(5.110) can be computed by a suitable PDE solver or by considering an accurate finite-dimensional approximation. The trajectories for $j = 1, 2, \ldots, N$ are finally obtained from

$$\tilde{x}^j_{n,i}(t) = \tilde{x}^{N+1}_{n,i}\left(t - T + (j-1)\tfrac{T}{N}\right), \qquad i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, N \tag{5.111}$$

$$\tilde{x}^j_{r,i}(t) = \tilde{x}^{N+1}_{r,i}\left(t - T + (j-1)\tfrac{T}{N}\right), \qquad i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, N \tag{5.112}$$

in view of (5.107)–(5.108) and the time-invariance of the considered systems, where $\tilde{x}^{N+1}_{n,i}(t) = \tilde{x}^{N+1}_{r,i}(t) = 0$ for $t < 0$. ◀

Note, that Remark 5.1-5 applies also here in an analog form. In the next subsection the closed-loop dynamics that result from the designed general sampling device and the observer-based compensator $\bar{\Sigma}^d_c$ (see (5.84)–(5.86)) are analyzed.

5.2.3 Analysis of the closed-loop dynamics

A state space model for the closed-loop system can be obtained by taking the system equations of $\bar{\Sigma}^d_n$, $\bar{\Sigma}^d_r$, and $\bar{\Sigma}^d_c$ into account (see (5.70)–(5.72), (5.77)–(5.79), and (5.84)–(5.86)). This yields

$$\bar{\Sigma}^d_{cl}: \quad \bar{x}_{cl}[k+1] = \bar{\mathcal{A}}_{d,cl}\bar{x}_{cl}[k], \qquad k \in \mathbb{N}_0, \quad \bar{x}_{cl}[0] = \bar{x}_{cl,0} \in \bar{X}_{cl} \tag{5.113}$$

with the state

$$\bar{x}_{cl} = \begin{bmatrix} \hat{\bar{x}}_n \\ \bar{x}_n \\ \bar{x}_r \end{bmatrix} \tag{5.114}$$

and the state space $\bar{X}_{cl} := \bar{X}_n \oplus \bar{X}_n \oplus \bar{X}_r$ (see (5.69) and (5.76)). The system operator in (5.113) has the decomposition $\bar{\mathcal{A}}_{d,cl} = \bar{\mathcal{A}}_{d,cl,0} + \bar{\Delta}$ with

$$\bar{\mathcal{A}}_{d,cl,0} = \begin{bmatrix} \bar{A}_{d,n} - L\bar{C}_{d,n} & 0 & 0 \\ \bar{B}_{d,n}K & \bar{A}_{d,n} - \bar{B}_{d,n}K & 0 \\ \bar{\mathcal{B}}_{d,r}K & -\bar{\mathcal{B}}_{d,r}K & \bar{\mathcal{A}}_{d,r} \end{bmatrix}, \qquad \bar{\Delta} = \begin{bmatrix} 0 & 0 & -L\bar{\mathcal{C}}_{d,r} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{5.115}$$

with $D(\bar{\mathcal{A}}_{d,cl,0}) = D(\bar{\Delta}) = \bar{X}_{cl}$. Comparison with (4.74)–(4.77) shows that the control loop considered here has the same structure as the closed-loop dynamics determined in Subsection 4.2.3. The only difference is that the state space X_{cl} is replaced here by the extended one $\bar{X}_{cl}$. For that reason, Corollary 4.2-2 needs to be modified slightly, at

which (5.83) has to be taken into account. This yields the following characterization of the closed-loop spectrum $\sigma(\bar{\mathcal{A}}_{d,cl})$.

Corollary 5.2-7

Let the Assumption 4.1-2 hold. Then, the spectrum $\sigma(\bar{\mathcal{A}}_{d,cl})$ of the closed-loop system operator in (5.113) can be decomposed as

$$\sigma(\bar{\mathcal{A}}_{d,cl}) = \{\tilde{\lambda}_{cl,i}, i \in \mathbb{N}\} \cup \sigma_c(\mathcal{A}_{d,r}) \cup \{\lambda_{d,0}\}, \tag{5.116}$$

where $\tilde{\lambda}_{cl,i}$, $i \in \mathbb{N}$, are eigenvalues of $\bar{\mathcal{A}}_{d,cl}$ that have finite algebraic multiplicities and are isolated, and $\lambda_{d,0} = 0$ is an eigenvalue with infinite multiplicity. Particularly,

$$\sigma_r(\bar{\mathcal{A}}_{d,cl}) = \emptyset \tag{5.117}$$

$$\sigma(\bar{\mathcal{A}}_{d,cl}) = \overline{\sigma_p(\bar{\mathcal{A}}_{d,cl})} \tag{5.118}$$

holds.

It has been argued before, that the use of the general sampling device leads to a small norm $\|\mathcal{C}_{d,r}^*\| \leq \|\mathcal{C}_{d,r}^*\|_{HS}$ if the sampling functions are designed by Corollary 5.2-5. Thus, the norm of the perturbation operator $\bar{\Delta}$ becomes small in view of

$$\|\bar{\Delta}\| = \|L\bar{\mathcal{C}}_{d,r}\| \leq \|L\|\|\mathcal{C}_{d,r}\| = \|L\|\|\mathcal{C}_{d,r}^*\| \tag{5.119}$$

(see (5.115), (5.82), and (5.87)). An expression for $\mathcal{C}_{d,r}^*$ follows from (5.95) when

$$\tilde{x}_{r,i}(t) = \sum_{j=1}^{N} \alpha_{j,i} \tilde{x}_{r,i}^j(t), \qquad i = 1, 2, \ldots, m \tag{5.120}$$

is taken into account, which follows from (5.93) by aid of $\tilde{\Pi} e_i(t) = \sum_{j=1}^{N} \alpha_{j,i} \kappa_j(t)$ and the definition of $\tilde{x}_{r,i}^j$. Thus, insertion into (5.95) yields

$$\mathcal{C}_{d,r}^* v = \sum_{j=1}^{N} \begin{bmatrix} \alpha_{j,1}\tilde{x}_{r,1}^j(T) & \alpha_{j,2}\tilde{x}_{r,2}^j(T) & \cdots & \alpha_{j,m}\tilde{x}_{r,m}^j(T) \end{bmatrix} v, \qquad \forall\, v \in \mathbb{C}^m. \tag{5.121}$$

By use of this and (5.119), Corollary 4.2-3 can be reformulated as follows.

Corollary 5.2-8

Let the Assumption 4.1-2 hold and assume that the observer-based compensator $\bar{\Sigma}_c^d$

is designed such that the eigenvalues of $\bar{A}_{d,n} - \bar{B}_{d,n}K$ *and* $\bar{A}_{d,n} - L\bar{C}_{d,n}$ *are simple, mutually different, and not contained in* $\sigma(\mathcal{A}_{d,r})$*. Then,*

$$d_d \leq \|\bar{\mathcal{T}}_{cl}^{-1}\|\|\bar{\mathcal{T}}_{cl}\|\|L\|\|\mathcal{C}_{d,r}^*\| \tag{5.122}$$

is an upper bound for the spectrum perturbation

$$d_d := \sup_{\tilde{\lambda}_{cl}\in\sigma(\bar{\mathcal{A}}_{d,cl})} \inf_{\lambda_{cl}\in\sigma(\bar{\mathcal{A}}_{d,cl,0})} |\tilde{\lambda}_{cl} - \lambda_{cl}|. \tag{5.123}$$

Therein, $\bar{\mathcal{T}}_{cl}$ *denotes a linear transformation such that* $\bar{\mathcal{T}}_{cl}^{-1}\bar{\mathcal{A}}_{d,cl,0}\bar{\mathcal{T}}_{cl}$ *is normal, and* $\mathcal{C}_{d,r}^*$ *is given in (5.121), where the* $\tilde{x}_{r,i}^j(T)$ *result from the approach of Theorem 5.2-5.*

Thus, the impact of the residual dynamics $\bar{\Sigma}_r^d$ on the closed-loop behavior can be suppressed by suitably choosing the sampling functions such that $\|\sum_{j=1}^N \alpha_{j,i}\tilde{x}_{r,i}^j(T)\|_{X_r}$ becomes small for $i = 1, 2, \ldots, m$. Note, that these norms can be made small just by choosing $C_{d,n}^{desired}$ in (5.104) with a small norm. This, however, leads to a large norm of L so that both effects cancel in the right hand-side of (5.122). Despite this, the spectrum perturbation d_d can be made arbitrarily small by choosing the number N of Dirac delta functions in (5.96)–(5.97) sufficiently large, as for the spillover reduction approach with a general hold device. This is stated next.

Theorem 5.2-9

Let the assumptions of Corollary 5.2-8 be satisfied. Then, for every $\varepsilon > 0$ *a number* N *of Dirac delta functions in (5.96)– (5.97) exists such that the design according to Theorem 5.2-5 yields* $d_d < \varepsilon$*.*

The proof is given in Appendix A.15. The approach of this section is demonstrated in the following example.

Example 5.2-10 (Control of an Euler-Bernoulli beam with Kelvin-Voigt damping, continued)

In Example 5.1-9 a sampled-data control with a general hold device was designed for an Euler-Bernoulli beam with Kelvin-Voigt damping. Instead of a general hold a general sampling device is designed now for the purpose of spillover reduction. The length $\ell = 1$ of the beam, the damping constant $\delta = 0.005$, and the input and output distribution

functions, given by (4.92), are considered as before, and it is again the intention to shift the most four dominant eigenvalues $\lambda_{\pm 1} = -0.487 \pm j9.86$ and $\lambda_{\pm 2} = -7.79 \pm j38.70$ of the continuous-time system operator $\mathcal{A}$. In view of the sampling constant $T = 0.025$ these assignments shall yield a stability margin $\beta_d = 0.839$ for the discrete-time closed-loop system. For the design of the compensator (5.84)–(5.86) the approximation from Example 4.2-6 is extended as in (5.70)–(5.75) which yields a discrete-time approximation $\bar{\Sigma}_n^d$ of order $n = 9$. Apparently, $\bar{C}_{d,n}$ depends on the sampling function whose design is described below. Note, that the dynamic matrix $\bar{A}_{d,n}$ of the approximation has the eigenvalues of $A_{d,n}$ and, in addition, an eigenvalue $\lambda_{d,0} = 0$ with algebraic multiplicity $\nu_0 = 5$ (see (5.73)). Since this latter eigenvalue is located already within the circle $\overline{\mathbb{C}}_{\beta_d}^i$ that corresponds to the specified stability margin $\beta_d = 0.839$, it needs not to be changed by the control. The eigenvalues of $A_{d,n}$ are shifted by assigning the eigenvalues $\lambda_{c,\pm 1} = -10 \pm j5$ and $\lambda_{c,\pm 2} = -10 \pm j10$ to the controlled continuous-time approximation as well as $\lambda_{o,\pm 1} = -15 \pm j5$ and $\lambda_{o,\pm 2} = -15 \pm j10$ to the continuous-time counterpart of the observer dynamics.

For the design of the general sampling device the approach of Theorem 5.2-5 and Remark 5.2-6 is applied. To this end, the state trajectory $\tilde{x}_{n,1}^{N+1}$ of the dual approximation (5.101) is computed, that results from the initial state $\tilde{x}_{n,1}^{N+1}(0) = C_n^*$ and the input $u(t) \equiv 0$ (see (5.109)). The trajectories $\tilde{x}_{n,1}^j$, $j = 1, 2, \ldots, N$, are obtained subsequently by suitable time-shifts of $\tilde{x}_{n,1}^{N+1}$ (see (5.111)). The trajectory $\tilde{x}_{r,1}^{N+1}$ of the dual residual dynamics (5.92) is computed on the basis of a modal approximation of order $n_r = 60$, where $\tilde{x}_{r,1}^{N+1}$ results from the initial state $\tilde{x}_{r,1}^{N+1}(0) = \mathcal{C}_r^*$ and the input $u(t) \equiv 0$ (see (5.110)). Afterwards, $\tilde{x}_{r,1}^j$, $j = 1, 2, \ldots, N$, is obtained by time-shifting $\tilde{x}_{r,1}^{N+1}$ in view of (5.112). In (5.104) the vector

$$C_{d,n}^{desired} := C_n = \begin{bmatrix} -0.0964 & -0.00976 & 0.0149 & 0.000377 \end{bmatrix} \tag{5.124}$$

is used. In this case it follows $\tilde{x}_{n,1}^1(T) = (C_n)^* = (C_{d,n}^{desired})^*$ from (5.109) and (5.111) so that (5.104) can be fulfilled for any $N \in \mathbb{N}$ at least by $\alpha_1 = [1 \ 0 \ \cdots \ 0]^T$. Furthermore, the eigenvalues of $A_{d,n}$, that shall be shifted by the compensator, are observable. The additional eigenvalue $\lambda_{d,0} = 0$ of $\bar{A}_{d,n}$ (see (5.73)) is unobservable, but is there is no need to shift it, as explained above. Since, in addition, $(\bar{A}_{d,n}, \bar{B}_{d,n})$ is controllable, the desired eigenvalues $\lambda_{d,c,i} = e^{\lambda_{c,i} T}$, $i = \pm 1, \pm 2$, can be assigned to $\bar{A}_{d,n} - \bar{B}_{d,n} K$ and $\lambda_{d,o,i} = e^{\lambda_{o,i} T}$ to $\bar{A}_{d,n} - L\bar{C}_{d,n}$. The minimization (5.102)–(5.104) finally yields the sampling function π_1 that is shown in Figure 35 for $N \in \{4, 10, 20\}$, and the corresponding closed-loop eigenvalue distributions are depicted in the Figures 36–38.

These figures make apparent that the perturbation of the desired eigenvalues becomes smaller when the number of Dirac delta impulses N is increased, as is expected in view of Theorem 5.2-9. Table 7 gives the related spectrum perturbations d_d. These have been determined on the basis of a modal approximation of the beam with order $n_{\text{high}} = 60$. For $N \in \{10, 20\}$ the desired stability margin $\beta_d = 0.839$ is achieved, as becomes apparent from the Figures 37–38. In summary, the use of a general sampling device yields a spillover reduction to qualitatively the same extent as the approach with a general hold device. However, the extension of the state spaces for the approach with general sampling leads to discrete-time approximations with the order $2n + m$, where n is the order of the corresponding continuous-time approximation. In contrast, the orders of the discrete-time and the continuous-time approximation coincide for the approach with a general hold device. ◀

Table 7 – Spectrum perturbation d_d of the closed-loop system with a general sampling device with $N = 4, 10, 20$ Dirac delta impulses. The case $N = 1$ corresponds to the standard sampling device. The Spectrum perturbation has been computated on the basis of a modal approximation of the beam with order $n_{\text{high}} = 60$.

N	1	4	10	20
d_d	0.329	0.173	0.108	0.0997

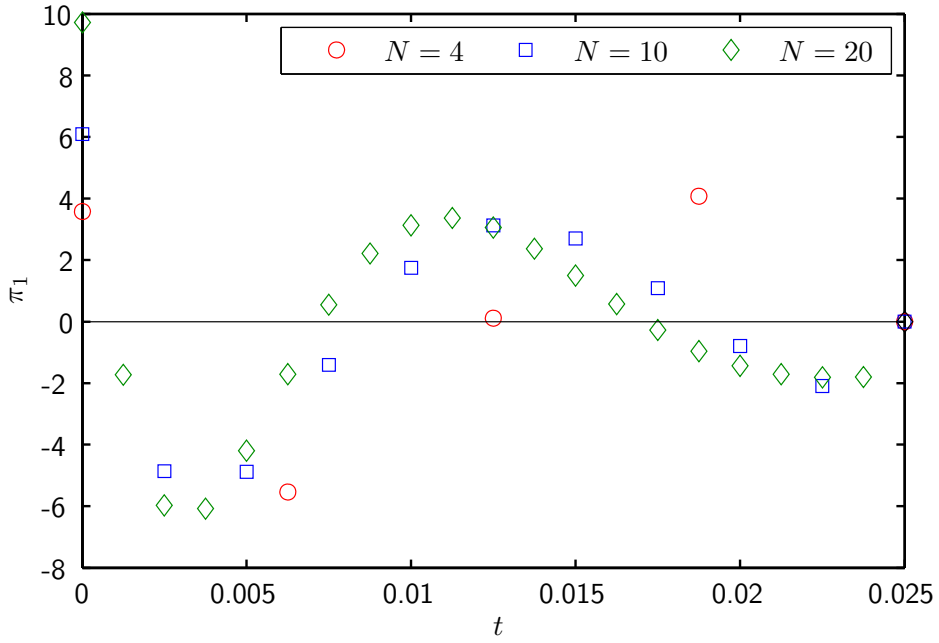

Figure 35 – Sampling functions with $N = 4, 10, 20$ Dirac delta impulses that minimize (5.102) subject to (5.104). A marker at (τ, α) stands for the Dirac delta function $\alpha\delta_\tau(t)$.

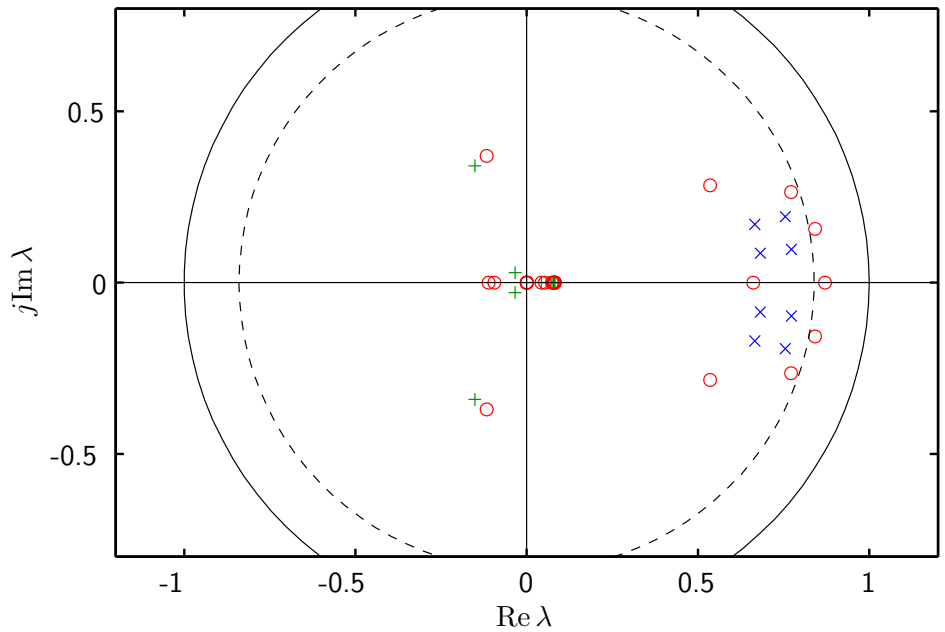

Figure 36 – Eigenvalues 'o' of the closed-loop system with a general sampling device with $N = 4$ steps. For comparison the desired eigenvalues are shown, where 'x' marks $\lambda_{d,c,\pm1}$, $\lambda_{d,c,\pm2}$, $\lambda_{d,o,\pm1}$, $\lambda_{d,o,\pm2}$, and '+' describes $\sigma(\mathcal{A}_{d,r})$. The dashed circle defines the boundary of the region $\overline{\mathbb{C}}^i_{\beta_d}$ that corresponds to the desired stability margin $\beta_d = 0.839$.

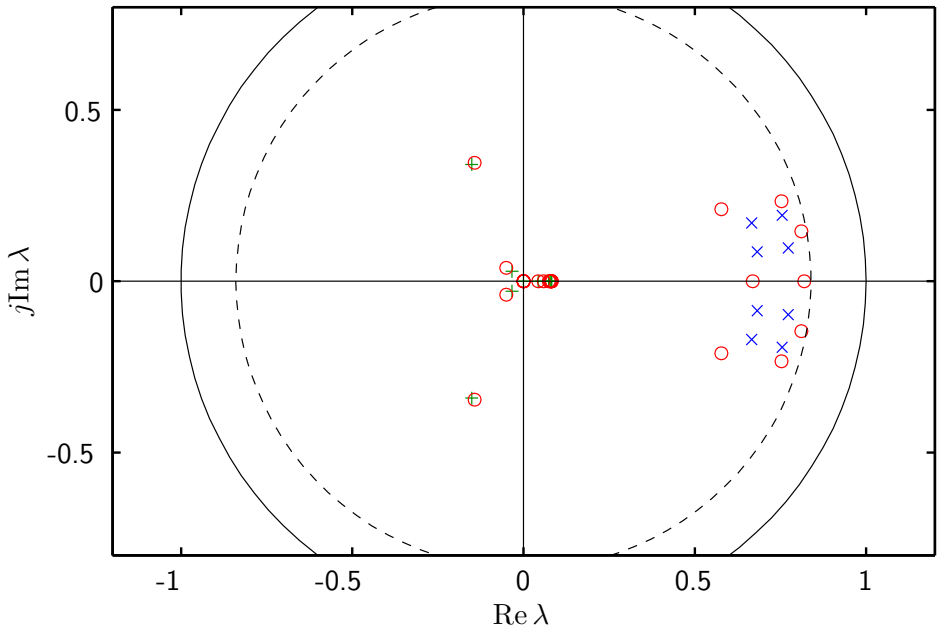

Figure 37 – Eigenvalues 'o' of the closed-loop system with a general sampling device with $N = 10$ steps. For comparison the desired eigenvalues are shown, where 'x' marks $\lambda_{d,c,\pm1}$, $\lambda_{d,c,\pm2}$, $\lambda_{d,o,\pm1}$, $\lambda_{d,o,\pm2}$, and '+' describes $\sigma(\mathcal{A}_{d,r})$.

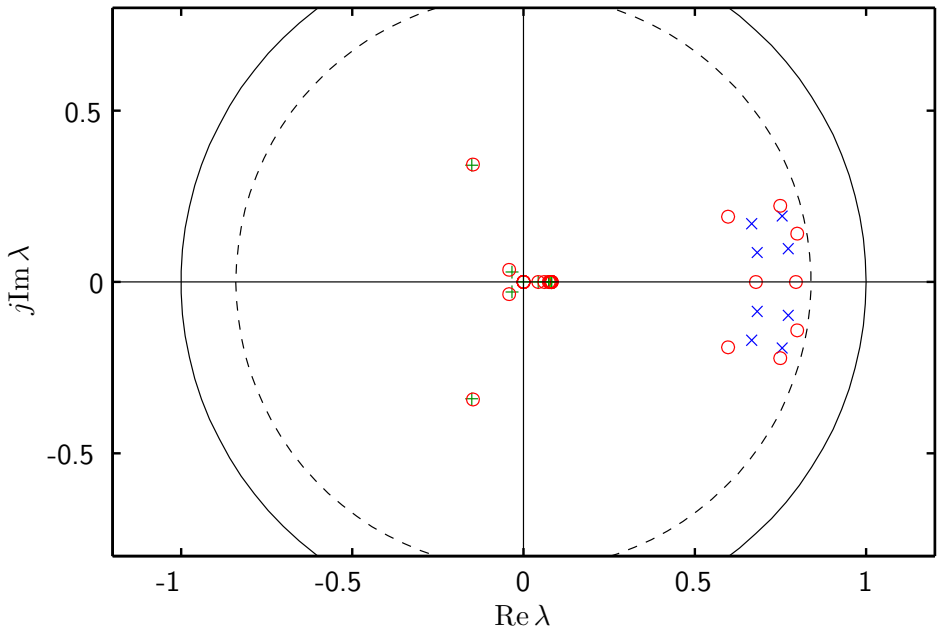

Figure 38 – Eigenvalues 'o' of the closed-loop system with a sampling hold device with $N = 20$ steps. For comparison the desired eigenvalues are shown, where 'x' marks $\lambda_{d,c,\pm1}$, $\lambda_{d,c,\pm2}$, $\lambda_{d,o,\pm1}$, $\lambda_{d,o,\pm2}$, and '+' describes $\sigma(\mathcal{A}_{d,r})$.

Chapter 6

Concluding remarks

The classical early-lumping based compensator design scheme exhibits two basic shortcomings: firstly, the spillover problem itself, that is inherently connected to utilizing an approximation in the overall approach. Secondly, the resulting design procedure, that consists of alternating compensator synthesis and control loop analysis steps, lacks an approach to improve the redesign such that the spillover is reduced systematically and efficiently. These problems are attenuated by the suggested approaches in the following way.

For the continuous-time control the impact of the neglected system dynamics is decreased by reconstructing suitable fictitious system outputs by output observers. Such an observer has the effect of a filter that suppresses the contributions of the residual dynamics to the system output but passes the contributions of the approximation. Repeated application of this filtering method allows to reduce the spillover in the sense that an upper bound for the provoked closed-loop spectrum perturbation is lowered. The implementation of the additional dynamics of the output observers can be considered as the expenses of the spillover reduction. However, since the perturbation estimate decreases exponentially with regard to the order of the additional dynamics the suggested approach is more effective than the bare increase of the approximation order. An a priori estimate is available that relates the order of the output observers to a maximum perturbation specification. Though being rather conservative this estimate helps to constitute a systematic design procedure. It should be remarked that the spillover reduction approach can be applied not only for the considered modal system approximations but also for alternative types of approximations. However, it is particularly simple to determine a state space model of the residual dynamics that correspond to a modal approximation and to analyze their spectrum.

For sampled-data systems two approaches for spillover reduction in the discrete-time domain are presented, where the system operator is assumed to be a Riesz-spectral operator. The first one makes use of a general hold device with specially shaped hold functions, and the second uses a general sampling device. The general sampling is designed such that the contribution of the residual dynamics to the output becomes small at the sampling time instances. The general hold, in contrast, is chosen such that the residual dynamics are excited only to a minor amount so that these again have small contributions to the plant output in consequence. Thus, the action of the digital control is affected only marginally for both approaches. Upper bounds for the perturbation of the closed-loop spectrum are provided for both methods. Since step functions are considered for the hold functions, it is possible to implement the general hold by a conventional zero-order hold that operates at a higher sampling rate in relation to the control. Similar, general sampling can be implemented by standard sampling with lower sampling constant because Dirac delta functions are used for the sampling functions. It has been shown in Section 3.3 that control spillover can be converted to observation spillover and vice versa. Therefore, it has the same spillover reducing effect, regardless whether a general hold device, that suppresses the control spillover, or general sampling, which aims at observation spillover reduction, is used. From the practical point of view, however, there are some aspects that help to decide between both approaches. Since the approach with the general hold device uses standard sampling and thus does not require measurements of the system output within the sampling intervals, this approach is more suitable when digital sensors with a comparatively low rate of measurements are used. In contrast, the approach with general sampling is more suitable when the actors cannot operate at the fast sampling rate of a general hold device. Another difference between both approaches is that a discrete-time approximation for general sampling has a higher order than a comparable discrete-time approximation for the general hold approach since the state space has to be extended in a certain way. Although the utilization of a general hold and general sampling have been discussed separately, it is possible to combine both approaches to yield an increased spillover suppression. Finally, if the additional efforts for implementing the general hold or the general sampling device are undesired, one has the possibility to adapt the spillover reduction approach for continuous-time control based on adding ouput observers to the control loop. Since this adaption to the discrete-time case is straightforward, it has not been discussed in the thesis.

In summary, both in the continuous-time and the discrete-time case the classical

early-lumping approach is extended to a systematic design approach for the finite-dimensional control of linear distributed-parameter systems.

The methods in the thesis have been presented under the assumption that the control and measurement is spatially distributed. However, the approaches can readily be applied to systems with boundary control by aid of the *boundary control system approach* in [46, Sec. 3.3], which enables to convert such an input to a distributed one. A desirable extension of the approaches, however, is the generalization to pointwise measurements and measurements at the boundary. A suitable framework to this end might be the theory of *regular linear systems* (see [130]).

An alternative to the observer-based compensators, that are considered throughout the thesis, is the use of a static output feedback controller for the plant that is dynamically extended by the output observers. This enables to assign some of the eigenvalues exactly, *i.e.*, without any spillover perturbation. This approach has been elaborated in [55, 56]. In addition, it is interesting to remark that the concept to reconstruct fictitious outputs has a dual counterpart that provides *fictitious inputs*. This is possible in the framework of the dual state feedback approach (see [50, 52]) that allows to assign some of the closed-loop eigenvalues without any spillover impact.

Besides in the field of infinite-dimensional systems the suggested methods can readily be applied for systems that are finite-dimensional but whose order is too high to design the compensator on the basis of the plant model. Systems with a very high but finite order result commonly from finite-elements models of distributed-parameter systems. In order to apply the presented compensator design approaches to such systems the high-dimensional state space of the plant plays the role of the infinite-dimensional state space considered throughout the thesis.

Appendix A

Proofs

A.1 Proof of Proposition 2.1-8

Due to Item 1 of Definition 2.1-7 there exist constants $a \in \mathbb{R}$, $\varepsilon \in (0, \frac{1}{2}\pi)$ such that $\sigma_p(\mathcal{A}) \subset S_{a,\varepsilon}$ (see (2.37)). Now, the operator

$$\mathcal{Q}h := \sum_{i=1}^{\infty} \frac{1}{\lambda_i - \lambda_0} \langle h, \psi_i \rangle_X \phi_i, \qquad \lambda_0 > a, \ \forall\, h \in X \tag{A.1}$$

is considered, in which ϕ_i are the eigenvectors of $\mathcal{A}$ that correspond to its eigenvalues λ_i and ψ_i are the eigenvectors of $\mathcal{A}^*$ corresponding to $\overline{\lambda_i}$. Note, that from $\sigma_p(\mathcal{A}) \subset S_{a,\varepsilon}$ it follows

$$|\lambda_i - \lambda_0| \geq \lambda_0 - a > 0, \qquad \forall\, i \in \mathbb{N} \tag{A.2}$$

so that the fraction in (A.1) is defined. Using (2.19) for (A.1) yields the upper bound

$$\|\mathcal{Q}h\|_X^2 \leq M \sum_{i=1}^{\infty} \left| \frac{1}{\lambda_i - \lambda_0} \langle h, \psi_i \rangle_X \right|^2 = M \sum_{i=1}^{\infty} \frac{1}{|\lambda_i - \lambda_0|^2} |\langle h, \psi_i \rangle_X|^2 \tag{A.3}$$

for $h \in X$, which by aid of (A.2) becomes

$$\|\mathcal{Q}h\|_X^2 \leq M \frac{1}{(\lambda_0 - a)^2} \sum_{i=1}^{\infty} |\langle h, \psi_i \rangle_X|^2 \tag{A.4}$$

for which the upper bound

$$M \frac{1}{(\lambda_0 - a)^2} \sum_{i=1}^{\infty} |\langle h, \psi_i \rangle_X|^2 \leq \frac{M}{m} \frac{1}{(\lambda_0 - a)^2} \Big\| \sum_{i=1}^{\infty} \langle h, \psi_i \rangle_X \phi_i \Big\|_X^2 = \frac{M}{m} \frac{1}{(\lambda_0 - a)^2} \|h\|_X^2 \tag{A.5}$$

again follows from (2.19) and (2.20). Thus, (A.4)–(A.5) show that $\mathcal{Q}$ is bounded and thus can be defined on whole X. From

$$\begin{aligned}(\mathcal{A}-\lambda_0 I)\mathcal{Q}h &= \mathcal{A}\sum_{i=1}^{\infty}\frac{1}{\lambda_i-\lambda_0}\langle h,\psi_i\rangle_X\phi_i - \lambda_0\sum_{i=1}^{\infty}\frac{1}{\lambda_i-\lambda_0}\langle h,\psi_i\rangle_X\phi_i\\ &= \sum_{i=1}^{\infty}\frac{\lambda_i}{\lambda_i-\lambda_0}\langle h,\psi_i\rangle_X\phi_i - \sum_{i=1}^{\infty}\frac{\lambda_0}{\lambda_i-\lambda_0}\langle h,\psi_i\rangle_X\phi_i\\ &= \sum_{i=1}^{\infty}\langle h,\psi_i\rangle_X\phi_i = h, \qquad \forall\, h\in X \end{aligned} \tag{A.6}$$

wherein (2.17) and (A.1) have been used, it follows that $\mathcal{Q}$ is the (bounded) inverse of $\mathcal{A}-\lambda_0 I$ that is defined on X. Thus, $\mathcal{A}-\lambda_0 I$ is closed according to [46, Thm. A.3.46] and hence also $\mathcal{A}$ is closed in view of [89, Problem III 5.6]. Taking the Items 2–4 of Definition 2.1-7 into account, all requirements of Definition 2.1-6 are satisfied so that $\mathcal{A}$ is a Riesz-spectral operator as stated.

That $\mathcal{A}$ is the infinitesimal generator of a C_0-semigroup if its eigenvalues λ_i satisfy the sector condition $\sigma_p(\mathcal{A})\subset S_{a,\varepsilon}$ for a sector angle $\varepsilon\geq 0$ follows from [46, Thm. 2.3.5]. Finally, it is stated in [46, Exercise 2.18] that $\mathcal{A}$ is the infinitesimal generator of an analytic C_0-semigroup if its eigenvalues λ_i satisfy the sector condition $\sigma_p(\mathcal{A})\subset S_{a,\varepsilon}$ for an $\varepsilon>0$. Thus, the proof is complete.

A.2 Proof of Proposition 2.2-3

First, observe that the biorthonormality relation $\langle\phi_i,\tilde{\psi}_j\rangle_X=\delta_{ij}$, $i,j=1,2,\ldots,n$ (see Assumption 2.2-2) implies that $\mathcal{P}$ according to (2.136) satisfies $\mathcal{P}^2h=\mathcal{P}h$, $\forall\, h\in X$, showing that this operator is a projection. Obviously, $\operatorname{ran}\mathcal{P}=\operatorname{span}\{\phi_i, i=1,2,\ldots,n\}=X_n$ and $\operatorname{nul}\mathcal{P}=\{h\in X\,|\,\langle h,\tilde{\psi}_i\rangle_X=0, i=1,2,\ldots,n\}=X_r$ hold which is why $\mathcal{P}$ projects onto X_n along X_r. In [106, Thm. 4.11.2] it is stated that these subspaces satisfy therefore $X_n+X_r=X$. Thus, the Items 1 and 2 of the proposition have been verified.

Now, it will be shown that X_n and X_r are $\mathcal{S}$-invariant. To this end, introduce the projection

$$\mathcal{P}_\Gamma h=\frac{1}{2\pi j}\int_\Gamma(\lambda I-\mathcal{A})^{-1}h\mathrm{d}\lambda, \qquad \forall\, h\in X, \tag{A.7}$$

in which Γ is the closed curve in the sense of Assumption 2.2-2. It follows from [46,

Lem. 2.5.7e] that this operator has the range

$$\operatorname{ran}\mathcal{P}_\Gamma = \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\} = X_n \tag{A.8}$$

when it is taken into account that λ_i, $1, 2, \ldots, n$, are assumed to have coinciding algebraic and geometric multiplicities. Using $\operatorname{nul}\mathcal{P}_\Gamma = (\operatorname{ran}\mathcal{P}_\Gamma^*)^\perp$ (see [89, Sec. III 3.3]) and

$$\mathcal{P}_\Gamma^* h = \frac{1}{2\pi j}\int_{\overline{\Gamma}} (\lambda I - \mathcal{A}^*)^{-1} h \mathrm{d}\lambda, \qquad \forall\, h \in X \tag{A.9}$$

(see [89, Sec. III 6.6]) one obtains

$$\operatorname{nul}\mathcal{P}_\Gamma = \left(\operatorname{ran}\frac{1}{2\pi j}\int_{\overline{\Gamma}} (\lambda I - \mathcal{A}^*)^{-1} h \mathrm{d}\lambda\right)^\perp. \tag{A.10}$$

Applying [46, Lem. 2.5.7e] again yields

$$\operatorname{nul}\mathcal{P}_\Gamma = (\operatorname{span}\{\psi_1, \psi_2, \ldots, \psi_n\})^\perp, \tag{A.11}$$

where ψ_i, $i = 1, 2, \ldots, n$, are the eigenvectors of $\mathcal{A}^*$ that correspond to $\overline{\lambda}_i$. Due to Assumption 2.2-2 the algebraic adjoint $\tilde{\mathcal{A}}^*$ has n eigenvectors $\tilde{\psi}_i$ that correspond to these eigenvalues, which are also eigenvectors of $\mathcal{A}^*$. Therefore, $\tilde{\psi}_i = \psi_i$, $i = 1, 2, \ldots, n$, has to hold. Thus, (A.11) gives $\operatorname{nul}\mathcal{P}_\Gamma = X_r$. This and (A.8) show that $\mathcal{P}_\Gamma$ is a projection onto X_n along X_r as is $\mathcal{P}$. However, since every $h \in X$ has a unique decomposition $h = h_n + h_r$ with $h_n \in X_n$ and $h_r \in X_r$ due to $X = X_n + X_r$ (see Item 1), the projection is uniquely determined so that $\mathcal{P}_\Gamma = \mathcal{P}$. It follows therefore from [46, Lem. 2.5.7a] that X_n and X_r are $\mathcal{S}$-invariant and from [46, Lem. 2.5.3] that X_n and X_r are $\mathcal{A}$-invariant. Thus, Item 3 of the proposition has been shown.

For verifying Item 4

$$\dot{\hat{x}}(t) = \mathcal{P}\dot{x}(t) = \mathcal{P}\mathcal{A}x(t) + \mathcal{P}\mathcal{B}u(t) = \mathcal{P}\mathcal{A}x_n(t) + \mathcal{P}\mathcal{A}x_r(t) + \mathcal{P}\mathcal{B}u(t) \tag{A.12}$$

is considered. Therein, one has $\mathcal{P}\mathcal{A}x_n(t) = \mathcal{A}x_n(t)$ because $\mathcal{A}x_n(t) \in X_n$ due to the $\mathcal{A}$-invariance of X_n and $\mathcal{P}h = h$, $\forall\, h \in X_n$. Furthermore, $\mathcal{P}\mathcal{A}x_r(t) = 0$ holds since $\mathcal{A}x_r(t) \in X_r$ and $\mathcal{P}h = 0$, $\forall\, h \in X_r$. Thus, (A.12) simplifies to (2.134). Similar,

$$\dot{x}_r(t) = \dot{x}(t) - \dot{x}_n(t) = \mathcal{A}x(t) + \mathcal{B}u(t) - \dot{x}_n(t) \tag{A.13}$$

simplifies to (2.135) by inserting (2.134), which completes the proof.

A.3 Proof of Theorem 2.3-3

In order to prove the theorem some auxiliary lemmas are presented first.

Lemma A.3-1

Let $\mathcal{D} : H \mapsto H$ denote a degenerate linear operator on a separable Hilbert space H with inner product $\langle \cdot, \cdot \rangle_H$. If $\mathcal{P}_N : H \mapsto H_N \subset H$ is a projection onto a finite-dimensional subspace H_N of H, then one has $\|\mathcal{D}(I - \mathcal{P}_N)\| \to 0$ for $N \to \infty$.

Proof. Let H_R denote an algebraic complement of H_N, *i.e.*, $H = H_N + H_R$. Since H is separable one has basis vectors φ_i, $i \in \mathbb{N}$, such that $H_N = \operatorname{span}\{\varphi_i, i = 1, 2, \ldots, N\}$ and $\{\varphi_i, i > N\}$ is a Riesz basis for H_R. Furthermore, one has corresponding biorthonormal vectors ψ_i, $i \in \mathbb{N}$, *i.e.*, $\langle \varphi_i, \psi_j \rangle_H = \delta_{ij}$, $i, j \in \mathbb{N}$. It is straightforward to check that

$$\mathcal{P}_N h = \sum_{i=1}^{N} \langle h, \psi_i \rangle_H \varphi_i, \qquad \forall\, h \in H \tag{A.14}$$

is a projection onto H_N along H_R, and $I - \mathcal{P}_N$ is given by

$$(I - \mathcal{P}_N)h = \sum_{i=N+1}^{\infty} \langle h, \psi_i \rangle_H \varphi_i, \qquad \forall\, h \in H. \tag{A.15}$$

Since $\mathcal{D}$ is degenerate it has the representation

$$\mathcal{D}h = \sum_{j=1}^{N_\mathcal{D}} \langle h, \gamma_j \rangle_H \zeta_j, \qquad \forall\, h \in H \tag{A.16}$$

with appropriate $\gamma_j \in H$, in which $\{\zeta_j, j = 1, 2, \ldots, N_\mathcal{D}\}$ is a basis for $\operatorname{ran} \mathcal{D}$ (see [89, Sec. III 4.3]). Thus, (A.15)–(A.16) yield

$$\mathcal{D}(I - \mathcal{P}_N)h = \sum_{j=1}^{N_\mathcal{D}} \sum_{i=N+1}^{\infty} \langle h, \psi_i \rangle_H \langle \varphi_i, \gamma_j \rangle_H \zeta_j, \qquad \forall\, h \in H. \tag{A.17}$$

Since $\{\zeta_j, j = 1, 2, \ldots, N_\mathcal{D}\}$ is a finite-dimensional basis for $\operatorname{ran} \mathcal{D}$ and thus a Riesz basis this leads to

$$\begin{aligned} \|\mathcal{D}(I - \mathcal{P}_N)h\|_H^2 &\leq M_\mathcal{D} \sum_{j=1}^{N_\mathcal{D}} \Big| \sum_{i=N+1}^{\infty} \langle h, \psi_i \rangle_H \langle \varphi_i, \gamma_j \rangle_H \Big|^2 \\ &\leq 2M_\mathcal{D} \sum_{j=1}^{N_\mathcal{D}} \sum_{i=N+1}^{\infty} |\langle h, \psi_i \rangle_H|^2 |\langle \varphi_i, \gamma_j \rangle_H|^2 \end{aligned} \tag{A.18}$$

with $M_{\mathcal{D}} > 0$ (see (2.19) and the parallelogram inequality). Since $\{\varphi_i, i > N\}$ is a Riesz basis for H_R also $\{\psi_i, i > N\}$ is a Riesz basis for H_R so that one has the representation

$$\gamma_j = \sum_{i=1}^{\infty} \langle \gamma_j, \varphi_i \rangle_H \psi_i, \qquad j = 1, 2, \ldots, N_{\mathcal{D}} \tag{A.19}$$

(see [46, Sec. 2.2]). Using again the Riesz basis property and (2.19) yields

$$\|\gamma_j\|_H^2 \geq m_j \sum_{i=1}^{\infty} |\langle \gamma_j, \varphi_i \rangle_H|^2, \qquad m_j > 0, \tag{A.20}$$

which implies $\sum_{i=N+1}^{\infty} |\langle \gamma_j, \varphi_i \rangle_H|^2 \to 0$ for $N \to \infty$. Using this in (A.18) shows $\|\mathcal{D}(I - \mathcal{P}_N)h\|_H \to 0$ for $N \to \infty$, $\forall\, h \in H$. Thus, $\|\mathcal{D}(I - \mathcal{P}_N)\| \to 0$ for $N \to \infty$ has been confirmed which completes the proof. ■

Lemma A.3-2

Let $\mathcal{M}_0 : D(\mathcal{M}_0) \subseteq H \mapsto H$ be a linear operator on a Hilbert space H. Suppose that $\mathcal{M}_0$ has eigenvectors ϕ_i, $i > N_0$, that, augmented by appropriate vectors $\varphi_i \in H$, $i = 1, 2, \ldots, N_0$, form a Riesz basis for H. Let $\mathcal{D} : H \mapsto H$ be a degenerate linear operator. Then, a number $N \geq N_0$ exists such that the following holds: $\mathcal{M} = \mathcal{M}_0 + \mathcal{D}$ has eigenvectors $\tilde{\phi}_i$, $i > N$, which can be augmented by appropriate vectors $\tilde{\varphi}_i \in H$, $i = 1, 2, \ldots, N$, such that $\{\tilde{\varphi}_i, i = 1, 2, \ldots, N\} \cup \{\tilde{\phi}_i, i > N\}$ forms a Riesz basis for H.

Proof. Define the subspaces

$$H_N \coloneqq \operatorname{span}(\{\varphi_i, i = 1, 2, \ldots, N_0\} \cup \{\phi_i, i = N_0 + 1, N_0 + 2, \ldots, N\}) \tag{A.21}$$

$$H_R \coloneqq \operatorname{span}\{\phi_i, i > N\} \tag{A.22}$$

with $N \geq N_0$, and let $\mathcal{P}_N : H \mapsto H_N$ denote the projection onto H_N along H_R. Thus, one has

$$\mathcal{P}_N \phi_i = 0, \qquad \forall\, i > N. \tag{A.23}$$

In the following, the eigenvectors of

$$\mathcal{M} = \mathcal{M}_0 + \mathcal{D} = \mathcal{M}_0 + \mathcal{D}\mathcal{P}_N + \mathcal{D}(I - \mathcal{P}_N) \tag{A.24}$$

are analyzed. Since H has a countable basis it is separable. Thus, Lemma A.3-1 can be applied, yielding $\|\mathcal{D}(I - \mathcal{P}_N)\| \to 0$ for $N \to \infty$. Since $\mathcal{D}$ is degenerate and thus bounded, this implies that $\mathcal{M}$ has eigenvectors $\tilde{\phi}_i$ for sufficiently large N that are

related to the eigenvectors of $\mathcal{M}_0 + \mathcal{D}\mathcal{P}_N$ by a linear transformation $\mathcal{T} : H \mapsto H$ (see [89, Thm. IV 2.24 and Thm. IV 4.16]). Note, that the eigenvectors ϕ_i, $i > N$, of $\mathcal{M}_0$ are also eigenvectors of $\mathcal{M}_0 + \mathcal{D}\mathcal{P}_N$ in view of (A.23). Thus, one has

$$\tilde{\phi}_i = \mathcal{T}\phi_i, \qquad i > N. \tag{A.25}$$

Since $\{\varphi_i, i = 1, 2, \ldots, N_0\} \cup \{\phi_i, i > N_0\}$ is a Riesz basis for H and $\mathcal{T}$ is a linear transformation also

$$\{\mathcal{T}\varphi_i, i = 1, 2, \ldots, N_0\} \cup \{\mathcal{T}\phi_i, i = N_0 + 1, N_0 + 2, \ldots, N\ \} \cup \{\tilde{\phi}_i, i > N\} \tag{A.26}$$

is a Riesz basis for H (see [46, Exercise 2.21]). This proves the assertion. ■

Lemma A.3-3

Let H_1 and H_2 be Hilbert spaces with inner product $\langle\cdot,\cdot\rangle_{H_1}$ and $\langle\cdot,\cdot\rangle_{H_2}$, respectively. Suppose $\mathcal{M}_{11} : D(\mathcal{M}_{11}) \subset H_1 \mapsto H_1$ and $\mathcal{M}_{22} : D(\mathcal{M}_{22}) \subset H_2 \mapsto H_2$ are the infinitesimal generators of C_0-semigroups on H_1 and H_2, respectively, that satisfy the SDGA, and $\mathcal{M}_{21} : H_1 \mapsto H_2$ is a bounded linear operator. Then,

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{11} & 0 \\ \mathcal{M}_{21} & \mathcal{M}_{22} \end{bmatrix} \tag{A.27}$$

is the infinitesimal generator of a C_0-semigroup on $H = H_1 \oplus H_2$ with the growth bound

$$\omega_0 = \sup_{\lambda \in \sigma(\mathcal{M}_{11}) \cup \sigma(\mathcal{M}_{22})} \operatorname{Re}\lambda \tag{A.28}$$

and satisfies the SDGA.

Proof. Since the SDGA holds for $\mathcal{M}_{ii}$, $i = 1, 2$, the corresponding C_0-semigroups $\mathcal{S}_i(t)$ satisfy

$$\|\mathcal{S}_i(t)\| \leq C_i e^{\omega_i t}, \qquad i = 1, 2,\ \forall\, t \geq 0 \tag{A.29}$$

for appropriate constants $C_i \geq 1$, depending on ω_i, and for all

$$\omega_i > \omega_{0,i} := \sup_{\lambda \in \sigma(\mathcal{M}_{ii})} \operatorname{Re}\lambda, \qquad i = 1, 2 \tag{A.30}$$

(see Subsection 2.1.3). In order to prove the assertion, that the C_0-semigroup $\mathcal{S}(t)$, generated by $\mathcal{M}$, has a growth bound ω_0 according to (A.28), it has to be shown that $\mathcal{S}(t)$ satisfies

$$\|\mathcal{S}(t)\| \leq C e^{\omega t}, \qquad \forall\, t \geq 0 \tag{A.31}$$

for all $\omega > \omega_0$ and some $C \geq 1$, depending on ω. Observe

$$\omega_0 = \max(\omega_{0,1}, \omega_{0,2}) \tag{A.32}$$

due to (A.28) and (A.30). For any $\tilde{\omega} > \omega_0$ choose $\omega_1 := \tilde{\omega}$ and $\omega_2 := \frac{1}{2}(\tilde{\omega} + \omega_0)$. Thus, $\omega_1 > \omega_0 \geq \omega_{0,1}$ and $\omega_2 > \omega_0 \geq \omega_{0,2}$ holds in view of (A.32), so that (A.29) holds. Furthermore, one has $\omega_1 \neq \omega_2$. Under these conditions [46, Lem. 3.2.2] can be applied, which states that the $\mathcal{S}(t)$ satisfies (A.31) for

$$\omega = \max(\omega_1, \omega_2) = \tilde{\omega} \tag{A.33}$$

Since $\tilde{\omega} > \omega_0$ is arbitrary (A.28) is confirmed. In order to show that $\mathcal{S}(t)$ satisfies the SDGA it remains to check that $\omega_0 = \sup_{\lambda \in \sigma(\mathcal{M})} \operatorname{Re} \lambda$ holds. This, however, is apparent from (A.28) and $\sigma(\mathcal{M}_{11}) \cup \sigma(\mathcal{M}_{22}) = \sigma(\mathcal{M})$, which comes from the block-triangle structure of $\mathcal{M}$. Thus, the proof is complete. ■

Lemma A.3-4

Let $\mathcal{M} : D(\mathcal{M}) \subseteq H \mapsto H$ be the infinitesimal generator of a C_0-semigroup on a Hilbert space H. Suppose that $\mathcal{M}$ has eigenvectors $\tilde{\phi}_i$, $i > N$, that, augmented by appropriate vectors $\tilde{\varphi}_i \in H$, $i = 1, 2, \ldots, N$, form a Riesz basis for H. Then, $\mathcal{M}$ is the infinitesimal generator of a C_0-semigroup that satisfies the SDGA.

Proof. Define

$$\tilde{H}_N := \operatorname{span}\{\tilde{\varphi}_i, i = 1, 2, \ldots, N\}, \qquad \tilde{H}_R := \operatorname{span}\{\tilde{\phi}_i, i > N\}. \tag{A.34}$$

Since $H = \tilde{H}_N + \tilde{H}_R$, the map

$$\iota : H \mapsto \tilde{H}_N \oplus \tilde{H}_R : h \mapsto \begin{bmatrix} h_N \\ h_R \end{bmatrix}, \quad h_N \in \tilde{H}_N,\, h_R \in \tilde{H}_R \tag{A.35}$$

is an isomorphism. One can check easily that the operator $\hat{\mathcal{M}} := \iota \mathcal{M} \iota^{-1}$ has the form

$$\hat{\mathcal{M}} = \begin{bmatrix} \hat{M}_{11} & 0 \\ \hat{\mathcal{M}}_{21} & \hat{\mathcal{M}}_{22} \end{bmatrix} \tag{A.36}$$

when it is taken into account that $\tilde{H}_R$ is $\mathcal{M}$-invariant. Since $\mathcal{M}$ is the infinitesimal generator of a C_0-semigroup by assumption $\hat{\mathcal{M}}_{22}$ is the infinitesimal generator of a C_0-semigroup in view of $\hat{\mathcal{M}}_{22} = \mathcal{M}|_{D(\mathcal{M}) \cap \tilde{H}_R}$. Furthermore, also $\hat{\mathcal{M}}_{11}$ is the infinitesimal generator of a C_0-semigroup since $\tilde{H}_N$ is finite-dimensional, and $\mathcal{M}_{21}$ is compact and

thus bounded for the same reason. Thus, Lemma A.3-3 can be applied, which states that $\hat{\mathcal{M}}$ and hence $\mathcal{M}$ satisfy the SDGA. So, the proof is complete. ■

Now, Theorem 2.3-3 can be proven easily by aid of the lemmas above. First, observe that $\mathcal{A}_{cl,0}$ has the eigenvectors

$$\phi_{cl,i} = \begin{bmatrix} 0 \\ 0 \\ \phi_{i-n} \end{bmatrix}, \qquad i > 2n, \tag{A.37}$$

in which ϕ_i, $i > n$, are eigenvectors of $\mathcal{A}_r$ and thus of $\mathcal{A}$ in view of the block-triangular structure of $\mathcal{A}_{cl,0}$ (see (2.177)). These eigenvectors are a Riesz basis for X_r due to (2.90) and in view of the assumption that $\mathcal{A}$ has an eigenvector Riesz basis for X. Thus, $\{\phi_{cl,i}, i > 2n\}$, augmented by any basis for $\mathbb{C}^n \oplus \mathbb{C}^n \oplus 0$, is a Riesz basis for X_{cl}. In addition, Δ is a degenerate operator (see (2.177)), so that Lemma A.3-2 can be applied to $\mathcal{A}_{cl,0}$ and Δ. It states that $\mathcal{A}_{cl} = \mathcal{A}_{cl,0} + \Delta$ has eigenvectors $\tilde{\phi}_i$, $i > N \geq 2n$, which can be augmented by appropriate vectors $\tilde{\varphi}_i \in X_{cl}$, $i = 1, 2, \ldots, N$, such that $\{\tilde{\varphi}_i, i = 1, 2, \ldots, N\} \cup \{\tilde{\phi}_i, i > N\}$ forms a Riesz basis for X_{cl}. Recall that $\mathcal{A}_{cl}$ has been found in Subsection 2.3.3 to be the infinitesimal generator of a C_0-semigroup on X_{cl}. Thus, Lemma A.3-4 can be applied to $\mathcal{A}_{cl}$ which states that this operator satisfies the SDGA, which completes the proof.

A.4 Proof of Lemma 2.4-1

For the proof the theory of *semi-Fredholm operators* is used. A closed, linear operator $\mathcal{F} : D(\mathcal{F}) \subset H \mapsto H$ is said to be semi-Fredholm if $\operatorname{ran} \mathcal{F}$ is closed and at least one of the spaces $\operatorname{nul} \mathcal{F}$ and $(\operatorname{ran} \mathcal{F})^\perp$ is finite-dimensional. Let $D_F(\mathcal{M}_0)$ denote the set of all $\lambda \in \mathbb{C}$ for which $\lambda I - \mathcal{M}_0$ is a semi-Fredholm operator. The complementary set of $D_F(\mathcal{F})$ with respect to $\mathbb{C}$ is introduced in [89, Sec. IV 5.6] as the *essential spectrum* $\sigma_e(\mathcal{M}_0) \coloneqq \mathbb{C} \setminus D_F(\mathcal{F}) \subseteq \sigma(\mathcal{F})$.

Now, Statement (S1) will be shown. In view of the general relation $\mathbb{C} = \rho(\cdot) \cup \sigma_p(\cdot) \cup \sigma_c(\cdot) \cup \sigma_r(\cdot)$ (see Subsection 2.1.3) and the assumption $\sigma_r(\mathcal{M}_0) = \emptyset$ (see Property P4) one has the decomposition

$$\mathbb{C} = \rho(\mathcal{M}_0) \cup \sigma_p(\mathcal{M}_0) \cup \sigma_c(\mathcal{M}_0), \tag{A.38}$$

where $\sigma_c(\mathcal{M}_0)$ contains the accumulation points of the eigenvalues of $\mathcal{M}_0$ if there are any. Due to Property P2 the eigenvalues of $\mathcal{M}_0$ are isolated and have finite multiplicities so that

$$\rho(\mathcal{M}_0) \cup \sigma_p(\mathcal{M}_0) \subseteq D_F(\mathcal{M}_0) \tag{A.39}$$

holds in view of [89, Thm. IV 5.28 and Problem IV 5.6]. In contrast,

$$\sigma_c(\mathcal{M}_0) \subseteq \sigma_e(\mathcal{M}_0) \tag{A.40}$$

follows from the fact that $\mathrm{ran}(\lambda I - \mathcal{M}_0)$ is open for any $\lambda \in \sigma_c(\mathcal{M}_0)$ (see [89, Thm. IV 5.2]) so that $\lambda I - \mathcal{M}_0$ cannot be semi-Fredholm according to the definition. Hence, (A.38)–(A.40) and $\mathbb{C} = D_F(\mathcal{M}) \cup \sigma_e(\mathcal{M})$ yield

$$\sigma_e(\mathcal{M}_0) = \sigma_c(\mathcal{M}_0). \tag{A.41}$$

It is an important result of the perturbation theory for semi-Fredholm operators that the essential spectrum is not changed under arbitrary compact[1] perturbations (see [89, Thm. IV 5.35]). Consequently, one has

$$\sigma_e(\mathcal{M}) = \sigma_e(\mathcal{M}_0) \tag{A.42}$$

since $\mathcal{D}$ is a degenerate operator and therefore compact. By aid of (A.41) one obtains $\sigma_e(\mathcal{M}) = \sigma_e(\mathcal{M}_0) = \sigma_c(\mathcal{M}_0)$, and since $\sigma_e(\mathcal{M}) \subset \sigma(\mathcal{M})$ this proves the part $\sigma_c(\mathcal{M}_0) \subset \sigma(\mathcal{M})$ in the statement. To verify the relation $\sigma_c(\mathcal{M}) \subseteq \sigma_c(\mathcal{M}_0)$ note that by an analog reasoning as for (A.40) it holds $\sigma_c(\mathcal{M}) \subseteq \sigma_e(\mathcal{M})$, and in virtue of (A.41)–(A.42) it follows $\sigma_c(\mathcal{M}) \subseteq \sigma_e(\mathcal{M}) = \sigma_e(\mathcal{M}_0) = \sigma_c(\mathcal{M}_0)$ as claimed.

Now, Statement (S3) shall be shown. From (A.41)–(A.42) it follows

$$\mathbb{C} \setminus \sigma_c(\mathcal{M}_0) = \mathbb{C} \setminus \sigma_e(\mathcal{M}) = D_F(\mathcal{M}), \tag{A.43}$$

where the right equality holds by definition. As argued in [89, Sec. IV 5.6] each of the spectral points of $\mathcal{M}$ within $D_F(\mathcal{M})$ is an isolated eigenvalue of $\mathcal{M}$ and has a finite algebraic multiplicity, provided that both of the following conditions are met:

1. $D_F(\mathcal{M})$ is *connected*[2] and

[1] A bounded linear operator $\mathcal{Q} : H_1 \mapsto H_2$ defined on whole H_1 is said to be *compact* if for any bounded sequence $(h_i)_{i\in\mathbb{N}}$ in H_1 the sequence $(\mathcal{Q}h_i)_{i\in\mathbb{N}}$ has a subsequence that is convergent in H_2. A sufficient condition for $\mathcal{Q}$ to be compact is that it is degenerate.

[2] A subset of a topological space is *connected* if it cannot be partitioned into two nonempty subsets such that each subset has no points in common with the set closure of the other.

2. $\rho(\mathcal{M}) \neq \emptyset$.

The first condition is satisfied because $D_F(\mathcal{M}) = \mathbb{C} \setminus \sigma_e(\mathcal{M})$ and $\sigma_e(\mathcal{M}) = \sigma_e(\mathcal{M}_0) \subset \sigma(\mathcal{M}_0)$ is totally disconnected (see Property P3). In order to verify the second condition the fact is used that, due to the boundedness of $\mathcal{D}$, the perturbed operator $\mathcal{M} = \mathcal{M}_0 + \mathcal{D}$ is the infinitesimal generator of a C_0-semigroup as $\mathcal{M}_0$ is (see Property P1 and [46, Thm. 3.2.1]) which implies $\rho(\mathcal{M}) \neq \emptyset$ (see [46, Lem. 2.1.11]). So, both conditions are met which proves Statement (S3) of the lemma.

Now, in order to show Statement (S2), assume that there exists a $\lambda \in \sigma_r(\mathcal{M})$. Then, $\dim \operatorname{nul}(\lambda I - \mathcal{M}) = 0$ holds because $\lambda I - \mathcal{M}$ is injective by the definition of $\sigma_r(\cdot)$, and $\operatorname{ran}(\lambda I - \mathcal{M})$ is closed which follows from the fact that $(\lambda I - \mathcal{M})^{-1}$ is bounded again due to the definition of $\sigma_r(\cdot)$ (see [89, Thm. IV 5.2]). Consequently, $\lambda \in D_F(\mathcal{M})$ follows by the properties of $D_F(\cdot)$. But in view of (A.43) this means $\lambda \in \mathbb{C} \setminus \sigma_c(\mathcal{M}_0)$ which contradicts Statement (S3) of the lemma.

Finally, the last statement can be shown by aid of the *Weinstein-Aronszajn determinant of the first kind*

$$\omega(\lambda) := \det((\mathcal{M} - \lambda I)(\mathcal{M}_0 - \lambda I)^{-1}) = \det(I + \mathcal{D}(\mathcal{M}_0 - \lambda I)^{-1}). \tag{A.44}$$

This function is analytic on any domain[3] of $\mathbb{C}$ that is a subset of $\rho(\mathcal{M}_0)$ and has (not removable) singularities at all spectral points of $\mathcal{M}_0$. To be more precise, if λ is a *pole*[4] of ω with order $\nu < \infty$, then λ is an eigenvalue of $\mathcal{M}_0$ with algebraic multiplicity ν. Similar, if λ is a zero of ω with order $\tilde{\nu}$ and $\lambda \in \rho(\mathcal{M}_0)$, then $\mathcal{M}$ has an eigenvalue with algebraic multiplicity $\tilde{\nu} < \infty$ (see [89, Sec. IV 6]). Now, consider a $\lambda \in \sigma_c^i(\mathcal{M})$ (if there is any), where $\sigma_c^i(\mathcal{M}) \subseteq \sigma_c(\mathcal{M})$ consists of all spectral points in $\sigma_c(\mathcal{M})$ that are isolated in $\sigma_c(\mathcal{M})$. Clearly, $\omega(\lambda)$ has a singularity at λ because $\lambda \in \sigma_c(\mathcal{M}) \subseteq \sigma_c(\mathcal{M}_0)$ (see Statement (S1)). This singularity is not a pole with finite order because λ is not an eigenvalue of $\sigma(\mathcal{M}_0)$ with finite multiplicity. Therefore, λ must be an essential singularity of ω. According to *Picard's Great Theorem* the map $\omega(\lambda)$ can have all complex values on any open set containing λ, with at most a single exception (see [33]). If the exceptional value η is different from zero, $\omega(\lambda)$ has therefore a zero in any neighborhood of λ. Taking into account that λ is an isolated spectral point of $\mathcal{M}_0$ so that λ has a neighborhood entirely belonging to $\rho(\mathcal{M}_0)$, this shows that $\mathcal{M}$ has

[3] A subset of a topological space is called a *domain* if it is nonempty, connected, and open.

[4] A singularity ω_0 of a function $f(\omega)$ is called a *pole with order* m if the Laurent series $f(\omega) = \sum_{k=-\infty}^{\infty} a_k(\omega - \omega_0)^k$ satisfies $a_{-m} \neq 0$ and $a_{-k} = 0$, $\forall\, k > m$.

eigenvalues arbitrarily close to λ. Thus, $\sigma_p(\mathcal{M})$ accumulates at λ which yields

$$\sigma_c^i(\mathcal{M}) \subseteq \overline{\sigma_p(\mathcal{M})}. \tag{A.45}$$

The case where the exceptional value η equals zero can be excluded as follows. In this case λ is not a point of accumulation but by replacing $\mathcal{D} = \mathcal{M} - \mathcal{M}_0$ by a modified perturbation $(1+\varepsilon)\mathcal{D}$ with $\varepsilon > 0$ arbitrarily small it follows $\eta \neq 0$ in view of the right part in (A.44). Thus, the property of λ being a point of accumulation depends discontinuously on ε which contradicts the fact that any isolated eigenvalue with finite multiplicity depends continuously on the perturbation $\mathcal{D}$ (see [89, Sec. IV 3.5]). In order to prove (2.191) it remains to consider the part $\sigma_c^a(\mathcal{M}) = \sigma_c(\mathcal{M}) \setminus \sigma_c^i(\mathcal{M})$. This set is totally disconnected as $\sigma(\mathcal{M}_0)$ is because of $\sigma_c^a(\mathcal{M}) \subseteq \sigma_c(\mathcal{M}) \subseteq \sigma_c(\mathcal{M}_0) \subset \sigma(\mathcal{M}_0)$ in view of Statement (S1), and because $\mathcal{M}_0$ is totally disconnected (see Property P3). For that reason and since the elements in $\sigma_c^a(\mathcal{M})$ are not isolated in $\sigma_c(\mathcal{M})$ by definition it follows that all elements of $\sigma_c^a(\mathcal{M})$ must be points of accumulation in $\sigma_c(\mathcal{M})$ which yields $\sigma_c^a(\mathcal{M}) \subset \overline{\sigma_c^i(\mathcal{M})}$ and thus

$$\sigma_c(\mathcal{M}) = \sigma_c^i(\mathcal{M}) \cup \sigma_c^a(\mathcal{M}) \subseteq \overline{\sigma_c^i(\mathcal{M})}. \tag{A.46}$$

Relation (A.45) implies $\overline{\sigma_c^i(\mathcal{M})} \subset \overline{\sigma_p(\mathcal{M})}$ so that by aid of (A.46) and $\sigma_r(\mathcal{M}) = \emptyset$ (see Statement (S2)) it has been shown that (2.191) is satisfied.

A.5 Proof of Lemma 2.4-6

First, it is assumed that $\mathcal{M}_0$ is normal so that $\mathcal{T} = I$ and hence (2.200) becomes

$$d_{\mathcal{M}} \leq \|\mathcal{D}\|, \tag{A.47}$$

which in view of (2.199) equivalently can be written as

$$\operatorname{dist}(\tilde{\lambda}, \sigma(\mathcal{M}_0)) \coloneqq \inf_{\lambda \in \sigma(\mathcal{M}_0)} |\tilde{\lambda} - \lambda| \leq \|\mathcal{D}\|, \qquad \forall\, \tilde{\lambda} \in \sigma(\mathcal{M}), \tag{A.48}$$

wherein $\operatorname{dist}(\omega, \Omega)$ denotes the *distance* between $\omega \in \mathbb{C}$ and $\Omega \subset \mathbb{C}$. The general case follows later. In case that $\sigma(\mathcal{M}) \subseteq \sigma(\mathcal{M}_0)$ (A.48) is trivially satisfied. Now, suppose $\sigma(\mathcal{M}) \not\subseteq \sigma(\mathcal{M}_0)$ so that there is a $\tilde{\lambda} \in \sigma(\mathcal{M}) \cap \rho(\mathcal{M}_0)$ in view of $\rho(\mathcal{M}_0) = \mathbb{C} \setminus \sigma(\mathcal{M}_0)$. As a first step it will be shown that such a $\tilde{\lambda}$ always satisfies $\|(\tilde{\lambda} I - \mathcal{M}_0)^{-1}\|_H^{-1} \leq \|\mathcal{D}\|_H$. To this end, consider a $\varphi \in D(\mathcal{M})$ with $\|\varphi\|_H = 1$ and define $\theta \coloneqq (\mathcal{M} - \tilde{\lambda} I)\varphi$, yielding

$$\tilde{\lambda}\, \varphi = \mathcal{M}\, \varphi - \theta, \qquad \tilde{\lambda} \in \sigma(\mathcal{M}) \cap \rho(\mathcal{M}_0). \tag{A.49}$$

Subtracting $\mathcal{M}_0\,\varphi$ on both sides of (A.49) and using $\mathcal{M}-\mathcal{M}_0=\mathcal{D}$ yields $(\tilde{\lambda}I-\mathcal{M}_0)\varphi=\mathcal{D}\,\varphi-\theta$ and thus

$$\varphi=(\tilde{\lambda}I-\mathcal{M}_0)^{-1}(\mathcal{D}\,\varphi-\theta) \tag{A.50}$$

because $\tilde{\lambda}\in\rho(\mathcal{M}_0)$. Taking the norm on both sides of (A.50) yields

$$1=\|(\tilde{\lambda}I-\mathcal{M}_0)^{-1}(\mathcal{D}\varphi-\theta)\|_H\leq\|(\tilde{\lambda}I-\mathcal{M}_0)^{-1}\|(\|\mathcal{D}\|+\|\theta\|_H), \tag{A.51}$$

wherein the triangular inequality and the fact that the operator norm $\|\cdot\|$ is submultiplicative are used. It is essential to observe, that $\|\theta\|_H$ can be made smaller than any positive number by a suitable choice of φ. To verify this for all $\tilde{\lambda}\in\sigma(\mathcal{M})\cap\rho(\mathcal{M}_0)$ it is sufficient to consider the two cases $\tilde{\lambda}\in\sigma_p(\mathcal{M})\cap\rho(\mathcal{M}_0)$ and $\tilde{\lambda}\in\sigma_c(\mathcal{M})\cap\rho(\mathcal{M}_0)$ because $\sigma_r(\mathcal{M})=\emptyset$ holds by assumption so that $\sigma(\mathcal{M})=\sigma_p(\mathcal{M})\cup\sigma_c(\mathcal{M})$. For the case $\tilde{\lambda}\in\sigma_p(\mathcal{M})\cap\rho(\mathcal{M}_0)$ take φ as an eigenvector of $\mathcal{M}$ corresponding to $\tilde{\lambda}$, so that $\theta=(\mathcal{M}-\tilde{\lambda}I)\varphi=0$ and thus $\|\theta\|_H=0$. For $\tilde{\lambda}\in\sigma_c(\mathcal{M})\cap\rho(\mathcal{M}_0)$ note that $(\tilde{\lambda}I-\mathcal{M})^{-1}$ is densely defined on H but unbounded. By the definition of the operator norm this means that for every $\alpha>0$ there is a $\psi\in H$ with $\|\psi\|_H\leq 1$ such that

$$\|(\mathcal{M}-\tilde{\lambda}I)^{-1}\psi\|_H>\alpha>0. \tag{A.52}$$

Choosing

$$\varphi=\frac{(\mathcal{M}-\tilde{\lambda}I)^{-1}\psi}{\|(\mathcal{M}-\tilde{\lambda}I)^{-1}\psi\|_H} \tag{A.53}$$

one obtains

$$\theta=(\mathcal{M}-\tilde{\lambda}I)\varphi=\frac{\psi}{\|(\mathcal{M}-\tilde{\lambda}I)^{-1}\psi\|_H}, \tag{A.54}$$

which yields

$$\|\theta\|_H\to 0\qquad\text{for}\qquad\alpha\to\infty \tag{A.55}$$

due to (A.52) and $\|\psi\|_H\leq 1$. Taking this result into account, (A.51) implies

$$\|(\tilde{\lambda}I-\mathcal{M}_0)^{-1}\|^{-1}\leq\|\mathcal{D}\|,\qquad\forall\tilde{\lambda}\in\sigma(\mathcal{M})\cap\rho(\mathcal{M}_0). \tag{A.56}$$

It is known that

$$\|(\tilde{\lambda}I-\mathcal{M}_0)^{-1}\|^{-1}=\inf_{\lambda\in\sigma(\mathcal{M}_0)}|\tilde{\lambda}-\lambda|=\operatorname{dist}(\tilde{\lambda},\sigma(\mathcal{M}_0)),\quad\forall\,\tilde{\lambda}\in\sigma(\mathcal{M})\cap\rho(\mathcal{M}_0) \tag{A.57}$$

holds for any normal operator $\mathcal{M}_0$ (see [89, Sec. V 8]). In addition, one has

$$\operatorname{dist}(\tilde{\lambda},\sigma(\mathcal{M}_0))=0,\qquad\forall\,\tilde{\lambda}\in\sigma(\mathcal{M})\cap\sigma(\mathcal{M}_0). \tag{A.58}$$

Thus, (A.56)–(A.58) yield (A.48).

Now, the case where $\mathcal{M}_0$ is not normal is considered. Applying the transformation $\mathcal{T}$ to the perturbed operator $\mathcal{M} = \mathcal{M}_0 + \mathcal{D}$ yields

$$\mathcal{T}^{-1}\mathcal{M}\mathcal{T} = \mathcal{T}^{-1}\mathcal{M}_0\mathcal{T} + \mathcal{T}^{-1}\mathcal{D}\mathcal{T}. \tag{A.59}$$

Therein, $\mathcal{T}^{-1}\mathcal{M}_0\mathcal{T}$ is normal by assumption and $\mathcal{T}^{-1}\Delta\mathcal{T}$ is bounded due to the boundedness of Δ, $\mathcal{T}$, and $\mathcal{T}^{-1}$ (see [89, Thm. III 4.8]). In addition, $\sigma_r(\mathcal{T}^{-1}\mathcal{M}\mathcal{T}) = \emptyset$ because $\sigma_r(\mathcal{M}) = \emptyset$ was assumed and

$$\sigma(\mathcal{T}^{-1}\mathcal{M}\mathcal{T}) = \sigma(\mathcal{M}) \tag{A.60}$$

holds due to the similarity of $\mathcal{T}^{-1}\mathcal{M}\mathcal{T}$ and $\mathcal{M}$. Consequently, $\mathcal{M}_0$, $\mathcal{M}$, and $\mathcal{D}$ in (A.48) can be replaced by $\mathcal{T}^{-1}\mathcal{M}_0\mathcal{T}$, $\mathcal{T}^{-1}\mathcal{M}\mathcal{T}$, and $\mathcal{T}^{-1}\mathcal{D}\mathcal{T}$, respectively, yielding

$$\operatorname{dist}(\tilde{\lambda}, \sigma(\mathcal{T}^{-1}\mathcal{M}_0\mathcal{T}))) \leq \|\mathcal{T}^{-1}\mathcal{D}\mathcal{T}\|, \qquad \forall\, \tilde{\lambda} \in \sigma(\mathcal{T}^{-1}\mathcal{M}\mathcal{T}). \tag{A.61}$$

Taking (A.60) and the analog relation $\sigma(\mathcal{T}^{-1}\mathcal{M}_0\mathcal{T}) = \sigma(\mathcal{M}_0)$ into account, (A.61) gives

$$\operatorname{dist}(\tilde{\lambda}, \sigma(\mathcal{M}_0)) \leq \|\mathcal{T}^{-1}\mathcal{D}\mathcal{T}\|, \qquad \forall\, \tilde{\lambda} \in \sigma(\mathcal{M}), \tag{A.62}$$

which is equivalent to (2.199)–(2.200).

A.6 Proof of Proposition 2.4-8

First, it is shown that $\mathcal{A}_{cl,0}$ is a Riesz-spectral operator under the assumptions in the proposition. For that, the four items in Definition 2.1-6 will be checked. In Section 2.3 it was argued that $\mathcal{A}_{cl,0}$ is the generator of a C_0-semigroup so that $\mathcal{A}_{cl,0}$ is closed (see [46, Thm. 2.1.10]) which is Item 1 of the definition. To verify the Items 2 and 3 first note that the eigenvalues of $\mathcal{A}_r$ are isolated and simple, and $\overline{\sigma_p(\mathcal{A}_r)}$ is totally disconnected in view of the assumed Riesz-spectral property of $\mathcal{A}$. Further, according to (2.179) $\sigma_p(\mathcal{A}_{cl,0})$ and $\sigma_p(\mathcal{A}_r)$ differ only by a finite number of eigenvalues which is why also $\overline{\sigma_p(\mathcal{A}_{cl,0})}$ is totally disconnected, and the contained eigenvalues are isolated and simple due to the prerequisites in the proposition. Thus, the Items 2 and 3 of the definition hold. Finally, in order to verify the Riesz basis property observe from

(2.177) that the eigenvectors $\phi_{cl,i}$ of $\mathcal{A}_{cl,0}$ have the form

$$\phi_{cl,i} = \begin{bmatrix} \phi_{2n,i} \\ \star \end{bmatrix}, \qquad i = 1, 2, \ldots, 2n \tag{A.63}$$

$$\phi_{cl,i} = \begin{bmatrix} 0 \\ \phi_{r,i-2n} \end{bmatrix}, \qquad i = 2n+1, 2n+2, \ldots \tag{A.64}$$

with $\phi_{2n,i}$ being the eigenvectors of the sub-block

$$A_{2n} = \begin{bmatrix} A_n - LC_n & 0 \\ B_nK & A_n - B_nK \end{bmatrix} \tag{A.65}$$

of $\mathcal{A}_{cl,0}$, and $\phi_{r,i}$ denoting the eigenvectors of $\mathcal{A}_r$. Since the eigenvalues of A_{2n} are simple by assumption, the vectors $\phi_{2n,i} \in \mathbb{C}^{2n}$, $i = 1, 2, \ldots, 2n$, are linearly independent and thus form a basis for $\mathbb{C}^{2n}$. In addition, the vectors $\phi_{r,i} \in X_r$, $i = 1, 2, \ldots$, constitute a Riesz basis for X_r because $\mathcal{A}_r$ inherits from $\mathcal{A}$ the property to be a Riesz-spectral operator. Finally, since the eigenvalues of A_{2n} are assumed to be disjoint from the eigenvalues of $\mathcal{A}_r$ so that the $\phi_{cl,i}$, $i = 1, 2, \ldots, 2n$, are linear independent from $\phi_{cl,i}$, $i = 2n+1, 2n+2, \ldots$ (see [92, Thm. 7.4-3]), this shows that $\{\phi_{cl,i}, i \in \mathbb{N}\}$ is a Riesz basis for $X_{cl} = \mathbb{C}^{2n} \oplus X_r$. In summary, $\mathcal{A}_{cl,0}$ is a Riesz-spectral operator. Therefore, the eigenvectors $\psi_{cl,i}$ of $\mathcal{A}_{cl,0}^*$ can be scaled such that $\{\phi_{cl,i}, i \in \mathbb{N}\}$ and $\{\psi_{cl,i}, i \in \mathbb{N}\}$ are biorthonormal sequences (see [46, Thm. 2.3.5]), as is assumed in the proposition. In consequence, it is easy to verify that $\mathcal{T}_{cl}^{-1}$ is given by (2.201). Taking the modal decomposition $\mathcal{A}_{cl,0}h = \sum_{i=1}^{\infty} \lambda_{cl,i} \langle h, \psi_i \rangle_{X_{cl}} \phi_{cl,i}$ and the biorthonormality of $\{\phi_{cl,i}, i \in \mathbb{N}\}$ and $\{\psi_{cl,i}, i \in \mathbb{N}\}$ into account, one obtains by aid of (2.201) that $\tilde{\mathcal{A}}_{cl,0} \coloneqq \mathcal{T}_{cl}^{-1} \mathcal{A}_{cl,0} \mathcal{T}_{cl}$ has the representation

$$\tilde{\mathcal{A}}_{cl,0}h = \sum_{i=1}^{\infty} \lambda_{cl,i} \langle h, \varphi_i \rangle_{X_{cl}} \varphi_i, \qquad h \in D(\tilde{\mathcal{A}}_{cl,0}) \tag{A.66}$$

which leads to

$$\tilde{\mathcal{A}}_{cl,0}^*h = \sum_{i=1}^{\infty} \overline{\lambda_{cl,i}} \langle h, \varphi_i \rangle_{X_{cl}} \varphi_i, \qquad h \in D(\tilde{\mathcal{A}}_{cl,0}^*). \tag{A.67}$$

Combining (A.66)–(A.67) yields

$$\tilde{\mathcal{A}}_{cl,0}\tilde{\mathcal{A}}_{cl,0}^*h = \sum_{i=1}^{\infty} \lambda_{cl,i}\overline{\lambda_{cl,i}} \langle h, \varphi_i \rangle_{X_{cl}} \varphi_i = \sum_{i=1}^{\infty} \overline{\lambda_{cl,i}}\lambda_{cl,i} \langle h, \varphi_i \rangle_{X_{cl}} \varphi_i = \tilde{\mathcal{A}}_{cl,0}^*\tilde{\mathcal{A}}_{cl,0}h, \tag{A.68}$$

where the orthonormality of $\{\varphi_i, i \in \mathbb{N}\}$ has been used. Moreover, using the Riesz basis property of $\{\phi_i, i \in \mathbb{N}\}$ and the orthonormality of $\{\varphi_i, i \in \mathbb{N}\}$ one can show that $\mathcal{T}_{cl}$ and $\mathcal{T}_{cl}^{-1}$ are defined on whole X which is why $\mathcal{T}_{cl}$ is a linear transformation. This

can be used to verify that $\tilde{\mathcal{A}}_{cl,0}$ is a Riesz-spectral operator as $\mathcal{A}_{cl,0}$ is. Therefore, one has the domains

$$D(\tilde{\mathcal{A}}_{cl,0}\tilde{\mathcal{A}}^*_{cl,0}) = \Big\{h \in X_{cl} \;\Big|\; \sum_{i=1}^{\infty} |\lambda_{cl,i}\overline{\lambda_{cl,i}}|^2 |\langle h, \varphi_i\rangle_{X_{cl}}|^2 < \infty\Big\} \tag{A.69}$$

$$D(\tilde{\mathcal{A}}^*_{cl,0}\tilde{\mathcal{A}}_{cl,0}) = \Big\{h \in X_{cl} \;\Big|\; \sum_{i=1}^{\infty} |\overline{\lambda_{cl,i}}\lambda_{cl,i}|^2 |\langle h, \varphi_i\rangle_{X_{cl}}|^2 < \infty\Big\} \tag{A.70}$$

(see [46, Thm. 2.3.5]) which reveals $D(\tilde{\mathcal{A}}_{cl,0}\tilde{\mathcal{A}}^*_{cl,0}) = D(\tilde{\mathcal{A}}^*_{cl,0}\tilde{\mathcal{A}}_{cl,0})$. This and (A.68) confirms that $\tilde{\mathcal{A}}_{cl,0} = \mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl}$ is normal (see (2.197)–(2.198)). Thus, the proof is complete.

A.7 Proof of Lemma 3.2-1

First, note that the operator $\mathcal{F}$ in (2.137) satisfies

$$\mathcal{F}v = \sum_{i=1}^{n} \langle v, e_i\rangle_{\mathbb{C}^n}\phi_i, \quad \forall\, v \in \mathbb{C}^n, \qquad \mathcal{F}^{-1}h = \sum_{i=1}^{n} \langle h, \tilde{\psi}_i\rangle_X e_i, \quad \forall\, h \in X_n, \tag{A.71}$$

where e_i denotes the i-th unit vector of $\mathbb{R}^n$ and $\langle \phi_i, \tilde{\psi}\rangle_X = \delta_{ij}$, $i, j = 1, 2, \ldots, n$, (see Assumption 2.2-2) is applied. Using this and the projection $\mathcal{P}$ according to (2.136), that yields $x_n(t) = \mathcal{F}^{-1}\mathcal{P}x(t)$ and $x_r(t) = (I - \mathcal{P})x(t)$, the fictitious output $y_o = \mathcal{C}(\mathcal{A} - \mu I)^{-1}x(t)$ (see (3.16) for $L_o = I$) can be written as

$$\begin{aligned} y_o(t) &= \mathcal{C}(\mathcal{A} - \mu I)^{-1}\mathcal{P}x(t) + \mathcal{C}(\mathcal{A} - \mu I)^{-1}(I - \mathcal{P})x(t) \\ &= \mathcal{C}\mathcal{F}\,\mathcal{F}^{-1}(\mathcal{A} - \mu I)^{-1}\mathcal{F}\,\mathcal{F}^{-1}\mathcal{P}x(t) + \mathcal{C}(\mathcal{A} - \mu I)^{-1}(I - \mathcal{P})x(t) \\ &= \mathcal{C}\mathcal{F}\,\mathcal{F}^{-1}(\mathcal{A} - \mu I)^{-1}\mathcal{F}\,x_n(t) + \mathcal{C}(\mathcal{A} - \mu I)^{-1}x_r(t). \end{aligned} \tag{A.72}$$

Taking $(\mathcal{A} - \mu I)^{-1}\phi_i = (\lambda_i - \mu)^{-1}\phi_i$ (see (2.17)) into account, one has in (A.72)

$$\mathcal{C}\mathcal{F}\,v = \sum_{i=1}^{n} \mathcal{C}\phi_i\langle v, e_i\rangle_{\mathbb{C}^n} = C_n v, \qquad \forall\, v \in \mathbb{C}^n \tag{A.73}$$

$$\begin{aligned} \mathcal{F}^{-1}(\mathcal{A} - \mu I)^{-1}\mathcal{F}\,v &= \sum_{i=1}^{n} \mathcal{F}^{-1}(\lambda_i - \mu)^{-1}\langle v, e_i\rangle_{\mathbb{C}^n}\phi_i \\ &= \sum_{i=1}^{n} (\lambda_i - \mu)^{-1}\langle v, e_i\rangle_{\mathbb{C}^n} e_i = (A_n - \mu I)^{-1}v, \quad \forall\, v \in \mathbb{C}^n, \end{aligned} \tag{A.74}$$

wherein (2.138) and (2.140) are used. Moreover, $(\mathcal{A} - \mu I)^{-1}$ in (A.72) can be replaced by $(\mathcal{A}_r - \mu I)^{-1}$ because it holds $\mathcal{A}_r = \mathcal{A}\big|_{X_r}$ (see (2.142)) and $x_r(t) \in X_r$. Similar,

$\mathcal{C}$ can be substituted by $\mathcal{C}_r$ since $\mathcal{C}_r = \mathcal{C}\big|_{X_r}$ (see (2.142)) and X_r is $\mathcal{A}$-invariant due to Proposition 2.2-3 implying $(\mathcal{A} - \mu I)^{-1} x_r(t) \in X_r$. Using these observations and (A.73)–(A.74) for (A.72) gives

$$y_o(t) = C_n(A_n - \mu I)^{-1} x_n(t) + \mathcal{C}_r(\mathcal{A}_r - \mu I)^{-1} x_r(t). \tag{A.75}$$

Finally, subtracting $e_o(t)$ on both sides of this equation yields

$$\hat{y}_o(t) = C_n(A_n - \mu I)^{-1} x_n(t) + \mathcal{C}_r(\mathcal{A}_r - \mu I)^{-1} x_r(t) - e_o(t) \tag{A.76}$$

in view of $y_o(t) - e_o(t) = \hat{y}_o(t)$ (see (3.4)). This confirms that (3.45) holds by aid of (3.43)–(3.44) and (3.48)–(3.49), which completes the proof.

A.8 Proof of Lemma 3.2-2

First, note that $T \coloneqq A_n - \mu I$ is invertible because $\mu \in S \subset \rho(\mathcal{A}) \subset \rho(A_n)$ is assumed (see (3.13) and (2.108)). Thus, T qualifies as a similarity transformation. Applying T to the realization $(C_{n,1}, A_n, B_n)$ of Σ_n^1 yields the new realization $(\tilde{C}_n, \tilde{A}_n, \tilde{B}_n)$ with

$$\tilde{C}_n = C_{n,1} T = C_{n,1}(A_n - \mu I) \tag{A.77}$$

$$\tilde{A}_n = T^{-1} A_n T = (A_n - \mu I)^{-1} A_n (A_n - \mu I). \tag{A.78}$$

Taking (3.44) into account as well as the fact that $(A_n - \mu I)^{-1}$ and A_n commute (see [89, Problem III 6.2]) (A.77)–(A.78) simplify to

$$\tilde{C}_n = C_n, \qquad \tilde{A}_n = A_n. \tag{A.79}$$

Since the properties detectability and observability are preserved under similarity transformations (see [86, Sec. 6.2.2]), $(C_{n,1}, A_n)$ is detectable, respectively observable, if and only if so is $(\tilde{C}_n, \tilde{A}_n)$, and in view of (A.79) this proves the statement.

A.9 Proof of Theorem 3.2-3

First,

$$\frac{\|\Delta_1\|}{\|\Delta\|} \leq \max(\eta_1, \eta_2^{-1}) \tag{A.80}$$

is shown instead of (3.67). From (3.66) it follows with $E := \{(h_1, h_2) \in \mathbb{C}^n \times X_r \mid \|h_1\|_{\mathbb{C}^n}^2 + \|h_2\|_{X_r}^2 \leq 1\}$ and the definition of the operator norm[5] that

$$\|\Lambda\| = \sup_{(h_1,h_2)\in E} \sqrt{\|(A_n - \mu I)h_1\|_{\mathbb{C}^n}^2 + \|(\mathcal{A}_r - \mu I)^{-1}h_2\|_{X_r}^2} \tag{A.81}$$

holds. The submultiplicativity of the operator norm applied to (3.65) yields $\|\Delta_1\| \leq \|\Delta\|\|\Lambda\|$ and insertion of (A.81) gives

$$\begin{aligned}
\|\Delta_1\| &\leq \|\Delta\| \sup_{(h_1,h_2)\in E} \sqrt{\|(A_n - \mu I)h_1\|_{\mathbb{C}^n}^2 + \|(\mathcal{A}_r - \mu I)^{-1}h_2\|_{X_r}^2} \\
&\leq \|\Delta\| \sup_{(h_1,h_2)\in E} \sqrt{\|A_n - \mu I\|^2\|h_1\|_{\mathbb{C}^n}^2 + \|(\mathcal{A}_r - \mu I)^{-1}\|^2\|h_2\|_{X_r}^2} \\
&\leq \|\Delta\| \max\left(\|A_n - \mu I\|, \|(\mathcal{A}_r - \mu I)^{-1}\|\right) \sup_{(h_1,h_2)\in E} \sqrt{\|h_1\|_{\mathbb{C}^n}^2 + \|h_2\|_{X_r}^2}.
\end{aligned} \tag{A.82}$$

Taking therein $\sup_{(h_1,h_2)\in E} \sqrt{\|h_1\|_{\mathbb{C}^n}^2 + \|h_2\|_{X_r}^2} = 1$ into account one obtains

$$\frac{\|\Delta_1\|}{\|\Delta\|} \leq \max\left(\|A_n - \mu I\|, \|(\mathcal{A}_r - \mu I)^{-1}\|\right) \tag{A.83}$$

which confirms (A.80) in view of (3.68).

Now, (3.67) shall be shown. To this end, it is essential to observe that η depends on the choice of the time coordinate. To explain this, the normalized time $\tau := t/T$ with respect to the *time basis* $T > 0$ is introduced. The dynamics of $\bar{x}(\tau) := x(t(\tau))$, given by

$$\frac{\mathrm{d}\bar{x}}{\mathrm{d}\tau}(\tau) = \frac{\mathrm{d}x}{\mathrm{d}t}(t(\tau))\,\frac{\mathrm{d}t}{\mathrm{d}\tau} = \big(\mathcal{A}x(t(\tau)) + \mathcal{B}u(t(\tau))\big)\,T = \mathcal{A}T\bar{x}(\tau) + \mathcal{B}T\bar{u}(\tau) \tag{A.84}$$

with $\bar{u}(\tau) := u(t(\tau))$, leads to the transformed plant model

$$\bar{\Sigma}: \quad \dot{\bar{x}}(\tau) = \bar{\mathcal{A}}\bar{x}(\tau) + \bar{\mathcal{B}}\bar{u}(\tau), \qquad \tau > 0, \quad \bar{x}(0) = x_0 \tag{A.85}$$

$$\bar{y}(\tau) = \mathcal{C}\bar{x}(\tau), \qquad \tau \geq 0 \tag{A.86}$$

with $\bar{\mathcal{A}} = \mathcal{A}T$, $\bar{\mathcal{B}} = \mathcal{B}T$, and $\bar{y}(\tau) := y(t(\tau))$. To analyze the influence of the time transformation on (A.80) observe that $\bar{A}_n = A_nT$ is the dynamic matrix of the approximation that corresponds to $\bar{\Sigma}$ (see (2.138)). Furthermore, the system operator $\bar{\mathcal{A}}_r$ of the related residual dynamics satisfies $\bar{\mathcal{A}}_r = \mathcal{A}_rT$ (see (2.142)), and the corresponding output observer has the eigenvalue $\bar{\mu} = \mu T$, which follows from reformulating (3.38)

[5] The *operator norm* $\|\mathcal{M}\|$ of a bounded operator $\mathcal{M}$ on a Banach space $(H, \|\cdot\|_H)$ is defined by $\|\mathcal{M}\| := \sup_{\|h\|_H \leq 1} \|\mathcal{M}h\|_H$.

w.r.t. the time coordinate τ. Applying the time transformation to the closed-loop system Σ_{cl} in (2.174) without output observer shows that its closed-loop system operator $\mathcal{A}_{cl}$ is modified to $\bar{\mathcal{A}}_{cl} = \mathcal{A}_{cl}T$, and the related perturbation operator is

$$\bar{\Delta} = \Delta T \tag{A.87}$$

(see (2.176)). Analog, the time transformation has the effect that the perturbation operator Δ_1 of the closed-loop system Σ_{cl}^1 in (3.59)–(3.60) with output observer is transformed to

$$\bar{\Delta}_1 = \Delta_1 T. \tag{A.88}$$

When the considerations for proving (A.80) are repeated on the basis of the state space models w.r.t. the new time coordinate τ one arrives at

$$\frac{\|\bar{\Delta}_1\|}{\|\bar{\Delta}\|} \leq \max(\bar{\eta}_1, \bar{\eta}_2^{-1}) \tag{A.89}$$

with

$$\bar{\eta}_1 = \|\bar{A}_n - \bar{\mu}I\| = \|A_n - \mu I\|T = \eta_1 T \tag{A.90}$$

$$\bar{\eta}_2^{-1} = \|(\bar{\mathcal{A}}_r - \bar{\mu}I)^{-1}\| = \|(\mathcal{A}_r - \mu I)^{-1}\|\frac{1}{T} = \eta_2^{-1}\frac{1}{T} \tag{A.91}$$

(see (3.68)). Inserting (A.87)–(A.88) and (A.90)–(A.91) into (A.89) yields

$$\frac{\|\Delta_1\|}{\|\Delta\|} \leq \max\left(\eta_1 T, \eta_2^{-1}\frac{1}{T}\right). \tag{A.92}$$

In order to obtain a tight estimate, the right hand-side is minimized over $T > 0$. Apparently, this minimum is yield if $\eta_1 T = \eta_2^{-1}/T$ holds, which gives

$$T = \frac{1}{\sqrt{\eta_1 \eta_2}}. \tag{A.93}$$

Inserting this into (A.92) yields (3.67), so that the proof is complete.

A.10 Proof of Theorem 3.3-1

In view of the norm $\|\cdot\|_{\tilde{X}_{cl}}$ defined in (3.137) it holds

$$\|\tilde{\mathcal{T}}h\|_{\tilde{X}_{cl}} = \|h\|_{X_{cl}}, \qquad \forall\, h \in X_{cl} \tag{A.94}$$

$$\|\tilde{\mathcal{T}}^{-1}h\|_{X_{cl}} = \|h\|_{\tilde{X}_{cl}}, \qquad \forall\, h \in \tilde{X}_{cl} \tag{A.95}$$

implying $\|\tilde{\mathcal{T}}\| = \|\tilde{\mathcal{T}}^{-1}\| = 1$. This shows that both operators $\tilde{\mathcal{T}} : X_{cl} \mapsto \tilde{X}_{cl}$ and $\tilde{\mathcal{T}}^{-1} : \tilde{X}_{cl} \mapsto X_{cl}$ are bounded which is why $\tilde{\mathcal{T}}$ is a bijection between X_{cl} and $\tilde{X}_{cl}$. Thus, $\tilde{x}_{cl}$ and x_{cl} can be converted into one another by (3.136). Let $\mathcal{S}_{cl}(t)$ denote the C_0-semigroup generated by $\mathcal{A}_{cl,q}$, *i.e.*, $x_{cl}(t) = \mathcal{S}_{cl}(t)x_{cl}(0)$ holds for $t \geq 0$. Using (3.136) this yields

$$\tilde{x}_{cl}(t) = \tilde{\mathcal{T}}x_{cl}(t) = \tilde{\mathcal{T}}\mathcal{S}_{cl}(t)\tilde{\mathcal{T}}^{-1}\tilde{x}_{cl}(0) = \tilde{\mathcal{S}}_{cl}(t)\tilde{x}_{cl}(0), \qquad t \geq 0 \tag{A.96}$$

with $\tilde{\mathcal{S}}_{cl}(t) = \tilde{\mathcal{T}}\mathcal{S}_{cl}(t)\tilde{\mathcal{T}}^{-1}$, $\forall\, t \geq 0$. This operator inherits the semigroup property from $\mathcal{S}_{cl}(t)$, *i.e.*, $\mathcal{S}_{cl}(t_1 + t_2) = \mathcal{S}_{cl}(t_1)\mathcal{S}_{cl}(t_2)$, $\forall\, t_1, t_2 \geq 0$, in view of

$$\begin{aligned}\tilde{\mathcal{S}}_{cl}(t_1 + t_2) &= \tilde{\mathcal{T}}\mathcal{S}_{cl}(t_1 + t_2)\tilde{\mathcal{T}}^{-1} = \tilde{\mathcal{T}}\mathcal{S}_{cl}(t_1)\mathcal{S}_{cl}(t_2)\tilde{\mathcal{T}}^{-1} \\ &= \tilde{\mathcal{T}}\mathcal{S}_{cl}(t_1)\tilde{\mathcal{T}}^{-1}\tilde{\mathcal{T}}\mathcal{S}_{cl}(t_2)\tilde{\mathcal{T}}^{-1} = \tilde{\mathcal{S}}_{cl}(t_1)\tilde{\mathcal{S}}_{cl}(t_2), \qquad \forall\, t_1, t_2 \geq 0.\end{aligned} \tag{A.97}$$

That $\tilde{\mathcal{S}}_{cl}(t)$ is in addition strongly continuous w.r.t. $\|\cdot\|_{\tilde{X}_{cl}}$ follows with (A.95) from

$$\begin{aligned}\|\tilde{\mathcal{S}}_{cl}(t)\tilde{x}_{cl}(0) - \tilde{x}_{cl}(0)\|_{\tilde{X}_{cl}} &= \|\tilde{\mathcal{T}}^{-1}\big(\tilde{\mathcal{T}}\mathcal{S}_{cl}(t)\tilde{\mathcal{T}}^{-1}\,\tilde{\mathcal{T}}x_{cl}(0) - \tilde{\mathcal{T}}x_{cl}(0)\big)\|_{X_{cl}} \\ &= \|\mathcal{S}_{cl}(t)x_{cl}(0) - x_{cl}(0)\|_{X_{cl}} \to 0 \quad \text{for} \quad t \to 0+\end{aligned} \tag{A.98}$$

for all $\tilde{x}_{cl}(0) \in \tilde{X}_{cl}$ since $\mathcal{S}_{cl}(t)$ is strongly continuous. Thus, $\tilde{\mathcal{A}}_{cl,q}$ is the infinitesimal generator of a C_0-semigroup as stated.

Now, $\sigma(\tilde{\mathcal{A}}_{cl,q}) = \sigma(\mathcal{A}_{cl,q})$ is proven by verifying $\rho(\tilde{\mathcal{A}}_{cl,q}) = \rho(\mathcal{A}_{cl,q})$, which is equivalent since the resolvent set and the spectrum are complementary sets in general. To this end, it is used that

$$(\lambda I - \mathcal{A}_{cl,q})g = h, \qquad h \in X_{cl} \tag{A.99}$$

has a unique solution g for any $h \in X_{cl}$ if and only if $\lambda \in \rho(\mathcal{A}_{cl,q})$ (see Subsection 2.1.3), which is assumed for the following. Taking (3.133) into account this equation can be rewritten as

$$(\lambda\tilde{\mathcal{T}}^{-1}\tilde{\mathcal{T}} - \tilde{\mathcal{T}}^{-1}\tilde{\mathcal{A}}_{cl,q}\tilde{\mathcal{T}})g = h, \qquad h \in X_{cl}, \tag{A.100}$$

which by multiplication with $\tilde{\mathcal{T}}$ from the left becomes

$$(\lambda\tilde{\mathcal{T}} - \tilde{\mathcal{A}}_{cl,q}\tilde{\mathcal{T}})g = (\lambda I - \tilde{\mathcal{A}}_{cl,q})\tilde{\mathcal{T}}g = \tilde{\mathcal{T}}h, \qquad h \in X_{cl}. \tag{A.101}$$

Since the right hand-side represents all elements of $\tilde{X}_{cl}$, and g and hence $\tilde{\mathcal{T}}g$ are uniquely determined this shows that λ is contained in $\rho(\tilde{\mathcal{A}}_{cl,q})$ whenever $\lambda \in \rho(\mathcal{A}_{cl,q})$. The same argumentation applies also in the reverse direction, which finally gives $\rho(\tilde{\mathcal{A}}_{cl,q}) = \rho(\mathcal{A}_{cl,q})$ and hence $\sigma(\tilde{\mathcal{A}}_{cl,q}) = \sigma(\mathcal{A}_{cl,q})$. Thus, the proof is complete.

A.11 Proof of Proposition 4.1-3

In order to prove that $\mathcal{A}_d$ is a sectorial Riesz-spectral operator the items of Definition 2.1-7 are checked. Comparison of (4.16) with the modal decomposition (2.33) of $\mathcal{A}$ shows that $\mathcal{A}_d$ has the same eigenvectors ϕ_i as $\mathcal{A}$, which hence form a Riesz basis for X because $\mathcal{A}$ is a Riesz-spectral operator by Assumption 4.1-2 (see Item 4 of Definition 2.1-6). Thus, Item 4 of the definition holds.

Taking the biorthonormality of ϕ_i and ψ_i into account (see (2.21)), (4.16) shows that the eigenvalue $\lambda_{d,i}$ of $\mathcal{A}_d$ that corresponds to ϕ_i can be expressed by a function $f : \mathbb{C} \mapsto \mathbb{C} \setminus \{0\}$ according to

$$\lambda_{d,i} = f(\lambda_i) = e^{\lambda_i T}, \qquad i \in \mathbb{N}, \tag{A.102}$$

wherein λ_i is an eigenvalues of $\mathcal{A}$. Note, that $\mathcal{A}_d$ cannot have any additional eigenvalue $\lambda_{d,0}$ besides the eigenvalues $\lambda_{d,i}$, $i \in \mathbb{N}$, because the eigenvector corresponding to $\lambda_{d,0}$ would be linear independent from ϕ_i, $i \in \mathbb{N}$, (see [92, Thm. 7.4-3]) which contradicts the fact that the latter ones form a Riesz basis for X. Thus, (4.33) is confirmed.

That the eigenvalues $\lambda_{d,i}$ satisfy the sector condition follows from the assumption that $\mathcal{A}$ is the infinitesimal generator of a C_0-semigroup (see Assumption 4.1-2) which implies $\sup_{i\in\mathbb{N}} \operatorname{Re}\lambda_i < \infty$ (see [46, Lem. 2.1.11]). Using (A.102) this shows that $\mathcal{A}_d$ has a bounded spectral radius $r_{\mathcal{A}_d} = \sup_{i\in\mathbb{N}} |\lambda_{d,i}|$ so that the sector condition holds. Thus, Item 1 of Definition 2.1-7 is confirmed.

In order to prove Item 3 assume that $\overline{\sigma_p(\mathcal{A}_d)}$ is not totally disconnected, *i.e.*, there exist two points $a, b \in \overline{\sigma_p(\mathcal{A}_d)}$ with $a \neq b$ that can be joined by a continuous open curve $\Gamma \subset \overline{\sigma_p(\mathcal{A}_d)}$. First, the situation is considered, where

$$\Gamma \subset F \coloneqq \mathbb{C} \setminus (-\infty, 0]. \tag{A.103}$$

Then, every $\gamma \in \Gamma$ is the limit point of a sequence $(\lambda_{d,i_k})_{k\in\mathbb{N}} \subset F$ because $\gamma \in \overline{\sigma_p(\mathcal{A}_d)}$. While the map f in (A.102) is not invertible on $\mathbb{C}$, it has an inverse $f^{-1} : F \mapsto \{s \in \mathbb{C} \,|\, -\pi < \operatorname{Im} s < \pi\}$ on the smaller set F, and one has

$$(f^{-1}(\lambda_{d,i_k}))_{k\in\mathbb{N}} = (\lambda_{i_k})_{k\in\mathbb{N}} \subset \sigma_p(\mathcal{A}) \tag{A.104}$$

due to (A.102). Since f^{-1} is continuous on F this sequence has an accumulation point $\tilde{\gamma} \in \overline{\sigma_p(\mathcal{A})}$, because $(\lambda_{d,i_k})_{k\in\mathbb{N}}$ accumulates at γ, and it holds $\tilde{\gamma} = f^{-1}(\gamma)$ (see [116, Thm. 4.6]). Using again the continuity of f^{-1} shows that the image

$$\tilde{\Gamma} \coloneqq \{f^{-1}(\gamma) \,|\, \gamma \in \Gamma\} \tag{A.105}$$

of Γ is a continuous curve contained in $\overline{\sigma_p(\mathcal{A})}$. This however contradicts the fact that $\overline{\sigma_p(\mathcal{A})}$ is totally disconnected since $\mathcal{A}$ is assumed to be a Riesz-spectral operator. Thus, a curve $\Gamma \subset F \cap \overline{\sigma_p(\mathcal{A}_d)}$ does not exist. Now, the case is considered, where Γ has one or several intersections with $(-\infty, 0]$ so that $\Gamma \not\subset F$. Then, Γ can be split at the intersection points into open curves that each are entirely contained in F. These, however, lead to the same contradiction as above which shows that $\overline{\sigma_p(\mathcal{A}_d)}$ is totally disconnected. This gives Item 3 of Definition 2.1-7.

Finally, Item 2 of Definition 2.1-7 has to be shown. To this end note, that f is injective on

$$E \coloneqq \{s \in \mathbb{C} \,|\, 0 \leq \operatorname{Im} s < 2\pi/T\} \tag{A.106}$$

for which reason the $\lambda_{d,i} = f(\lambda_i)$ for $\lambda_i \in E$ are simple and mutual different because the eigenvalues λ_i are simple and mutual different due to the Riesz-spectral property of $\mathcal{A}$. Relation (4.32) assures that the same is true for the images of all eigenvalues of $\mathcal{A}$ so that $\lambda_{d,i}$, $i \in \mathbb{N}$, are distinct numbers. It follows from [46, Cor. 2.3.6] that $\mathcal{A}_d$ is therefore a Riesz-spectral operator when it is taken into account that it has the representation (4.16), its eigenvectors ϕ_i form a Riesz basis for X, and $(\phi_i)_{i\in\mathbb{N}}$ and $(\psi_i)_{i\in\mathbb{N}}$ are biorthonormal sequences. Thus, Item 2 of Definition 2.1-7 is verified in view of Item 2 of Definition 2.1-6 so that, in summary, $\mathcal{A}_d$ is a sectorial Riesz-spectral operator.

The equations (4.35)–(4.36) immediately follow from (2.34)–(2.35) since $\mathcal{A}$ and $\mathcal{A}_d$ are Riesz-spectral operators. The equality in (4.34) is obtained from $\sigma(\mathcal{A}_d) = \sigma_p(\mathcal{A}_d) \cup \sigma_c(\mathcal{A}_d) \cup \sigma_r(\mathcal{A}_d)$ and (4.35)–(4.36). Finally, the inclusion in (4.34) follows from (4.26), (4.33), and (4.35). Thus, the proof is complete.

A.12 Proof of Theorem 5.1-3

First, it is shown that the Items 1 and 2 of Problem 5.1-2 hold. Item 1 is apparently satisfied due to the ansatz $\theta_1(t) = \sum_{j=1}^{N} \alpha_j \kappa_j(t)$ and in view of (5.27). Since (5.31)–(5.32) and (5.35) yield

$$B_{d,n} = x_{n,1}(T) = \sum_{j=1}^{N} \alpha_j x_{n,1}^j(T) = B_{d,n}^{desired} \tag{A.107}$$

also Item 2 is confirmed. Finally, it has to be shown that $\|\mathcal{B}_{d,r}\|$ is minimized in the case of single-input systems for $\theta_1(t) = \sum_{j=1}^{N} \alpha_j \kappa_j(t)$. To this end, the norm of the

operator

$$\mathcal{B}_{d,r}v = x_{r,1}(T)v = \sum_{j=1}^{N} \alpha_j x_{r,1}^j(T)\, v, \qquad \forall\, v \in \mathbb{C} \tag{A.108}$$

(see (5.26) and (5.30)) is expressed by

$$\begin{aligned} \|\mathcal{B}_{d,r}\|^2 &= \left\langle \sum_{j=1}^{N} \alpha_j x_{r,1}^j(T), \sum_{k=1}^{N} \alpha_k x_{r,1}^k(T) \right\rangle_{X_r} \\ &= \sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_j \alpha_k \langle x_{r,1}^j(T), x_{r,1}^k(T)\rangle_{X_r} = \alpha^T \tilde{M} \alpha \end{aligned} \tag{A.109}$$

with α as defined in (5.34) and

$$\tilde{M} = \begin{bmatrix} \langle x_{r,1}^1(T), x_{r,1}^1(T)\rangle_{X_r} & \cdots & \langle x_{r,1}^1(T), x_{r,1}^N(T)\rangle_{X_r} \\ \vdots & & \vdots \\ \langle x_{r,1}^N(T), x_{r,1}^1(T)\rangle_{X_r} & \cdots & \langle x_{r,1}^N(T), x_{r,1}^N(T)\rangle_{X_r} \end{bmatrix}, \tag{A.110}$$

where it is used that the α_j are real. Note, that $\tilde{M}$ is hermitian. It can be shown easily that for this reason $\operatorname{Im}\tilde{M}$ is skew-symmetric, and that $\operatorname{Im}\tilde{M}$ does therefore not contribute to the right hand-side in (A.109). Consequently, $\tilde{M}$ therein can be replaced by $M = \operatorname{Re}\tilde{M}$. Thus, if the objective function $\alpha^T M \alpha$ is minimized, as claimed in (5.33), then $\|\mathcal{B}_{d,r}\|$ becomes minimal. So, Problem 5.1-2 is satisfied, which completes the proof.

A.13 Proof of Theorem 5.1-8

Recall that $x_{r,i}(T)$, $i = 1, 2, \ldots, p$, is the state of the continuous-time residual dynamics Σ_r at the instant of time $t = T$ for vanishing initial state and input $u(t) = e_i\theta_i(t) = \sum_{j=1}^{N} e_i \alpha_{j,i} \kappa_j(t)$ (see (5.24) and (5.45)). By introducing the *controllability map* $\mathcal{B}_r^t : L_2([0,t];\mathbb{R}^p) \mapsto X_r$, defined by

$$\mathcal{B}_r^t u := \int_0^t \mathcal{S}_r(t-\tau)\mathcal{B}_r u(\tau)\mathrm{d}\tau, \qquad \forall\, u \in L_2([0,t];\mathbb{R}^p), \tag{A.111}$$

this can be expressed by

$$x_{r,i}(T) = \mathcal{B}_r^T e_i \theta_i, \tag{A.112}$$

wherein (5.25) and $\Theta e_i = e_i\theta_i$ are used. Analog, $x_{n,i}$, $i = 1, 2, \ldots, p$, is the state of the continuous-time approximation Σ_n for vanishing initial state and input $u(t) = e_i\theta_i(t)$

which can be written as $x_{n,i}(T) = \mathcal{B}_n^T e_i \theta_i$ with

$$\mathcal{B}_n^t u := \int_0^t e^{A_n(t-\tau)} B_n u(\tau) \mathrm{d}\tau, \qquad \forall\, u \in L_2([0,t];\mathbb{R}^p) \tag{A.113}$$

(see [86, Sec. 2.5]). In the following, input signals $\tilde{u} \in E_i$ with

$$E_i := \left\{ \tilde{u} = e_i \tilde{\theta}_i \,|\, \tilde{\theta}_i \in L_2([0,T];\mathbb{R}),\, \mathcal{B}_n^T e_i \tilde{\theta}_i = B_{d,n}^{desired} e_i \right\}, \qquad i = 1,2,\ldots,p \tag{A.114}$$

are considered. The hold functions $\tilde{\theta}_i$ that define this set are not confined to the set K_Θ of step functions (see (5.27)–(5.28)) but assure that Item 2 of Problem 5.1-2 is satisfied (see (5.31)).

For the moment it is assumed that

$$\overline{\mathrm{ran}(\mathcal{B}_r^T|_{E_i})} = X_r, \qquad i = 1,2,\ldots,p \tag{A.115}$$

is satisfied which means that the space of input signals $u \in E_i$ allows to approach all states of the residual dynamics arbitrarily close. Thus, a feedforward control $\tilde{u}(t) = e_i \tilde{\theta}_i(t) \in E_i$ exists such that the state

$$\tilde{x}_{r,i}(T) = \mathcal{B}_r^T \tilde{u} \tag{A.116}$$

of Σ_r, that corresponds to the input $\tilde{u}$ and vanishing initial state, satisfies $\|\tilde{x}_{r,i}(T)\| < \varepsilon$ for arbitrary $\varepsilon > 0$. Since the hold functions θ_i are confined to the set K_Θ of step functions, the control $\tilde{u}(t) = e_i \tilde{\theta}_i(t)$ may not be implementable by the general hold device exactly. However, the error $\|\tilde{u} - u\|_{L_2([0,T];\mathbb{R}^p)}$ can be made arbitrarily small by using a sufficiently large number N of steps, which follows from the construction of the Lebesgue integral and the L_2 space (see, *e.g.*, [112, Sec. 2.3]). Thus, $\|\tilde{u} - u\|_{L_2([0,T];\mathbb{R}^p)} < \tilde{\varepsilon}$ for arbitrary $\tilde{\varepsilon} > 0$ may be assumed. Taking into account that $\mathcal{B}_r^T$ is bounded (see [46, Lem. 4.1.4]) it follows by aid of (A.112) and (A.116)

$$\|x_{r,i}(T) - \tilde{x}_{r,i}(T)\|_{X_r} = \|\mathcal{B}_r^T (u - \tilde{u})\|_{X_r} \leq \|\mathcal{B}_r^T\| \tilde{\varepsilon}. \tag{A.117}$$

Using the triangular inequality yields

$$\|x_{r,i}(T)\|_{X_r} \leq \|\tilde{x}_{r,i}(T)\|_{X_r} + \|x_{r,i}(T) - \tilde{x}_{r,i}(T)\|_{X_r} \leq \varepsilon + \|\mathcal{B}_r^T\| \tilde{\varepsilon}. \tag{A.118}$$

Since $\varepsilon > 0$ and $\tilde{\varepsilon} > 0$ can be made arbitrarily small as argued above, this shows that for a sufficiently large N a hold function $\theta_i \in K_\Theta$ exists such that $\mathcal{B}_{d,r} e_i = x_{r,i}(T)$ (see (5.26)) satisfies

$$\|\mathcal{B}_{d,r} e_i\|_{X_r} < \delta, \qquad i = 1,2,\ldots,p \tag{A.119}$$

for any given $\delta > 0$. Since the approach of Theorem 5.1-3 and Remark 5.1-6 minimizes $\|\mathcal{B}_{d,r} e_i\|$, this approach yields θ_i such that (A.119) indeed holds if N is sufficiently large. Thus,

$$\|\mathcal{B}_{d,r}\| = \left\| \begin{bmatrix} x_{r,1}(T) & x_{r,2}(T) & \cdots & x_{r,p}(T) \end{bmatrix} \right\| \to 0 \quad \text{for} \quad N \to \infty. \tag{A.120}$$

Hence, by help of (5.51) the statement is proven under the assumption (A.115). Note, that the norms $\|\tilde{\mathcal{T}}_{cl}^{-1}\|$, $\|\tilde{\mathcal{T}}_{cl}\|$, and $\|K\|$ in (5.51) are independent from N because only $\mathcal{B}_{d,r}$ depends on the choice of the general sampling device when Item 2 of Problem 5.1-2 is taken into account.

Now, the situation is considered, where (A.115) does not hold. To this end, the state space X_r is divided into $X_r = \tilde{X}_r + \tilde{X}_r^{\perp}$, where

$$\tilde{X}_r = \overline{\operatorname{ran}(\mathcal{B}_r^T|_{E_i})}. \tag{A.121}$$

By restricting Σ_r to the subspace $\tilde{X}_r$, *i.e.*, X_r, $\mathcal{A}_r$, and $\mathcal{C}_r$ in (5.17)–(5.18) are replaced by $\tilde{X}_r$, $\mathcal{A}_r|_{\tilde{X}_r}$, and $\mathcal{C}_r|_{\tilde{X}_r}$, respectively, the resulting system $\tilde{\Sigma}_r$ is approximately controllable (see [123, Thm. 9.1.9]). Hence, when the state $x_{r,i}$ is decomposed according to

$$x_{r,i}(t) = x_{r,i}^{\parallel}(t) + x_{r,i}^{\perp}(t), \qquad x_{r,i}^{\parallel}(t) \in \tilde{X}_r,\ x_{r,i}^{\perp}(t) \in \tilde{X}_r^{\perp} \tag{A.122}$$

the same considerations as above can be applied to the part $x_{r,i}^{\parallel}$, yielding

$$\|x_{r,i}^{\parallel}(T)\|_{X_r} \to 0 \quad \text{for} \quad N \to \infty. \tag{A.123}$$

The part $x_{r,i}^{\perp}$ satisfies

$$x_{r,i}^{\perp}(T) = 0 \tag{A.124}$$

because one has $x_{r,i}(T) \in \tilde{X}_r$ due to $x_{r,i}(0) = 0$ and the definition of $\tilde{X}_r$. Thus, (A.122)–(A.124) and $\mathcal{B}_r e_i = x_{r,i}(T)$ yields (A.120). In view of (5.51), this finally confirms the statement so that the proof is complete.

A.14 Proof of Proposition 5.2-1

Let $y^{x[k-1]}(t)$ denote the output of the plant Σ in (5.58)–(5.59) for $t \in [t_{k-1}, t_k)$, $k \in \mathbb{N}$, that results from $u \equiv 0$ and $x[k-1] \in X$, and similar, let $y^u(t)$ be the output of Σ for $t \in [t_{k-1}, t_k)$, $k \in \mathbb{N}$, that results from $x[k-1] = 0$ and $u \not\equiv 0$. Then, the output y of Σ can be decomposed into

$$y(t) = y^{x[k-1]}(t) + y^u(t), \qquad t \in [t_{k-1}, t_k),\ k \in \mathbb{N} \tag{A.125}$$

due to the linearity of Σ. Using (2.120) the part $y^{x[k-1]}$ in turn has the decomposition

$$y^{x[k-1]}(t) = y_n^{x[k-1]}(t) + y_r^{x[k-1]}(t), \qquad t \in [t_{k-1}, t_k),\ k \in \mathbb{N}, \tag{A.126}$$

where $y_n^{x[k-1]}$ denotes the output of the continuous-time approximation Σ_n in (4.48)–(4.49) with $u \equiv 0$, and $y_r^{x[k-1]}$ is the output of the continuous-time residual dynamics Σ_r in (4.56)–(4.57) for $u \equiv 0$. By taking $y_d[k] = \int_0^T \Pi(\tau) y(\tau + t_{k-1}) \mathrm{d}\tau$, $k \in \mathbb{N}$, (see (5.57)) and (A.125)–(A.126) into account, the discrete-time output y_d can be written as

$$y_d[k] = \int_0^T \Pi(\tau) y_n^{x[k-1]}(\tau + t_{k-1}) \mathrm{d}\tau + \int_0^T \Pi(\tau) y_r^{x[k-1]}(\tau + t_{k-1}) \mathrm{d}\tau + \int_0^T \Pi(\tau) y^u(\tau + t_{k-1}) \mathrm{d}\tau \tag{A.127}$$

for $k \in \mathbb{N}$, and $y_d[0] = 0$. Now, observe that $y_n^{x[k-1]}$ is obtained from (4.48)–(4.49) and $x_n(t_{k-1}) = x_n[k-1]$ by

$$y_n^{x[k-1]}(\tau + t_{k-1}) = C_n x_n(\tau + t_{k-1}) = C_n e^{A_n \tau} x_n[k-1] \tag{A.128}$$

for $\tau \in [0, T)$ (see [86, Sec. 2.5]), and $y_r^{x[k-1]}$ satisfies analog

$$y_r^{x[k-1]}(\tau + t_{k-1}) = \mathcal{C}_r x_r(\tau + t_{k-1}) = \mathcal{C}_r \mathcal{S}_r(\tau) x_r[k-1] \tag{A.129}$$

in view of (4.56)–(4.57) and $x_r(t_{k-1}) = x_r[k-1]$ (compare to (2.13)). Furthermore, (5.58)–(5.59) yields

$$y^u(\tau + t_{k-1}) = \mathcal{C} x(\tau + t_{k-1}) = \mathcal{C} \int_0^\tau \mathcal{S}(\kappa) \mathcal{B} u(\tau + t_{k-1} - \kappa) \mathrm{d}\kappa = \mathcal{C}\mathcal{B}_d(\tau) u_d[k-1] \tag{A.130}$$

for $\tau \in [0, T)$, where (5.60) and (5.61) have been used (see [46, Def. 3.1.4, Thm. 3.1.7]). Insertion of (A.128)–(A.130) into (A.127) confirms that $y_{d,n}$ and $y_{d,r}$ according to (5.63) and (5.65) satisfy $y_d[k] = y_{d,n}[k] + y_{d,r}[k]$, as claimed. Finally, the equations (5.62) and (5.64) have been determined in Subsection 4.2.1 (see (4.50) and (4.58)). Thus, the proof is complete.

A.15 Proof of Theorem 5.2-9

Consider some sampling functions $\hat{\pi}_i \in L_2([0, T]; \mathbb{R})$, $i = 1, 2, \ldots, m$, that are not constraint to consist of Dirac delta functions but are such that Item 2 of Problem 5.2-4 is satisfied. It can be shown in the same way as in the proof of Theorem 5.1-8 (see Appendix A.13) that these sampling functions can be chosen such that the norms

$\|\hat{x}_{r,i}(T)\|_{X_r}$ become arbitrarily small, where $\hat{x}_{r,i}$ is the state of the dual residual dynamics (5.92) for the input $u(t) = e_i\hat{\pi}_i(T-t)$ and $\hat{x}_{r,i}(0) = 0$. For the operator

$$\hat{\mathcal{C}}_{d,r}h := \int_0^T \operatorname{diag}(\hat{\pi}_1(\tau), \hat{\pi}_2(\tau), \ldots, \hat{\pi}_m(\tau))\, \mathcal{C}_r\mathcal{S}_r(\tau)\mathrm{d}\tau\, h, \qquad \forall\, h \in X_r, \tag{A.131}$$

that results when the sampling functions $\hat{\pi}_i$ are used in (5.87) instead of π_i, one has $\hat{\mathcal{C}}_{d,r}^* e_i = \hat{x}_{r,i}(T)$, $i = 1, 2, \ldots, m$, analog to (5.95). Consequently, the above reasoning shows that one can assume

$$\|\hat{\mathcal{C}}_{d,r}^*\|_{HS} < \hat{\varepsilon} \tag{A.132}$$

for arbitrary $\hat{\varepsilon} > 0$. The corresponding general sampling law (5.54) with π_i replaced by $\hat{\pi}_i$ cannot be implemented exactly in general since the sampling functions π_i are confined to the set K_Π of weighted Dirac delta functions, which means that π_i has the form $\pi_i(t) = \sum_{j=1}^N \alpha_{j,i}\delta_{(j-1)\frac{T}{N}}(t)$ (see (5.96)–(5.97)), so that the integrals in (5.54) become for $k = 1$

$$\int_0^T \pi_i(\tau)y_i(\tau)\mathrm{d}\tau = \sum_{j=1}^N \alpha_{j,i}\, y_i\big((j-1)\tfrac{T}{N}\big). \tag{A.133}$$

In order to make the error between this sum and the integral $\int_0^T \hat{\pi}_i(\tau)y_i(\tau)\mathrm{d}\tau$ small, that is the analog to the left hand-side of (A.133) for the use of the sampling function $\hat{\pi}_i$, this integral is approximated by the *rectangle method* for integration calculus, *i.e.*,

$$\int_0^T \hat{\pi}_i(\tau)y_i(\tau)\mathrm{d}\tau \approx \frac{T}{N}\sum_{j=1}^N \hat{\pi}_i\big((j-1)\tfrac{T}{N}\big)y_i\big((j-1)\tfrac{T}{N}\big) \tag{A.134}$$

(see, *e.g.*, [74, Sec. I.3]). Apparently, by choosing $\alpha_{j,i} = \frac{T}{N}\hat{\pi}_i\big((j-1)\frac{T}{N}\big)$ the right hand-sides of (A.133)–(A.134) coincide, which gives

$$\sum_{j=1}^N \alpha_{j,i}\, y_i\big((j-1)\tfrac{T}{N}\big) \approx \int_0^T \hat{\pi}_i(\tau)y_i(\tau)\mathrm{d}\tau \tag{A.135}$$

and equivalently

$$\sum_{j=1}^N \operatorname{diag}\big(\alpha_{j,1}, \ldots, \alpha_{j,m}\big)\, y\big((j-1)\tfrac{T}{N}\big) \approx \int_0^T \operatorname{diag}\big(\hat{\pi}_1(\tau), \ldots, \hat{\pi}_m(\tau)\big)y(\tau)\,\mathrm{d}\tau. \tag{A.136}$$

By substituting $y(t)$ therein by $\mathcal{C}_r\mathcal{S}_r(t)h$ and taking (A.131) and

$$\begin{aligned}\sum_{j=1}^N \operatorname{diag}\big(\alpha_{j,1}, \ldots, \alpha_{j,m}\big)\, \mathcal{C}_r\mathcal{S}_r\big((j-1)\tfrac{T}{N}\big)h \\ = \int_0^T \Pi(\tau)\mathcal{C}_r\mathcal{S}_r(\tau)\mathrm{d}\tau\, h = \mathcal{C}_{d,r}h, \qquad \forall\, h \in X_r\end{aligned} \tag{A.137}$$

(see (5.56), (5.87), and (5.98)) into account, (A.136) becomes

$$\mathcal{C}_{d,r}h \approx \hat{\mathcal{C}}_{d,r}h, \qquad \forall\, h \in X_r. \tag{A.138}$$

It is well-known that the error in (A.134) and thus also in (A.136) can be made arbitrarily small for sufficiently large N if y is continuous (see [74, Sec. I.3]). For (A.137)–(A.138) $y(t)$ has been replaced by $y_r^{x[0]}(t) \coloneqq \mathcal{C}_r\mathcal{S}_r(t)h$ which is the output of the residual dynamics Σ_r that results from $u \equiv 0$ and $x_r(0) = h$ (see (A.129)). This output is known to be continuous (see [46, Lem. 3.1.5]). Therefore,

$$\|\mathcal{C}_{d,r}^* - \hat{\mathcal{C}}_{d,r}^*\|_{HS} = \|\mathcal{C}_{d,r} - \hat{\mathcal{C}}_{d,r}\|_{HS} < \varepsilon, \qquad \varepsilon > 0 \tag{A.139}$$

is achieved if N is sufficiently large. Taking (A.132) into account, this shows that for sufficiently large N sampling functions $\pi_i \in K_\Pi$ exist such that

$$\|\mathcal{C}_{d,r}^*\|_{HS} \leq \|\mathcal{C}_{d,r}^* - \hat{\mathcal{C}}_{d,r}^*\|_{HS} + \|\hat{\mathcal{C}}_{d,r}^*\|_{HS} < \varepsilon + \hat{\varepsilon} \tag{A.140}$$

by the triangular equation. Since the approach of Theorem 5.2-5 minimizes $\|\mathcal{C}_{d,r}^*\|_{HS}$, this approach yields π_i such that (A.140) indeed holds if N is sufficiently large. Thus, by help of (5.122) and $\|\mathcal{C}_{d,r}^*\| \leq \|\mathcal{C}_{d,r}^*\|_{HS}$ the statement is proven since the norms $\|\bar{\mathcal{T}}_{cl}^{-1}\|$, $\|\bar{\mathcal{T}}_{cl}\|$, and $\|L\|$ in (5.122) can be verified to be independent from N by use of Item 2 of Problem 5.2-4.

Appendix B

Computation of the disk radius for spectrum enclosure

In this appendix the numerical computation of the disk radius r for the spectrum enclosure (2.204)–(2.205) is addressed. Indeed, making use of the fact that $\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}$ has finite rank, *i.e.*, $\dim \operatorname{ran}(\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}) < \infty$, it is possible to convert the computation of $\|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\|$ to the standard problem of determining the norm of a finite-dimensional matrix. The following lemma provides a step-by-step procedure for the calculation of r that is suitable for the computation by standard numeric software. However, it has to be assumed that $\mathcal{A}$ is normal. An important class of operators that satisfies this condition consists of Riesz-spectral operators with orthonormal eigenvectors $\phi_i, i \in \mathbb{N}$.

Theorem B-1

Suppose that $\mathcal{A}$ is normal. Accomplish the following procedure:

1. *Compute the eigenvectors $\phi_{cl,i}$, $i \leq 2n$, of the operator $\mathcal{A}_{cl,0}$ by solving the boundary value problems*

$$\mathcal{A}_{cl,0}\,\phi_{cl,i} = \lambda_{cl,i}\phi_{cl,i}, \quad \phi_{cl,i} \in D(\mathcal{A}_{cl,0}), \quad i = 1, 2, \ldots, 2n \tag{B.1}$$

   for the assigned eigenvalues $\lambda_{cl,i} \in \sigma(A_n - B_nK) \cup \sigma(A_n - LC_n)$.

2. *Define the $2n \times 2n$-matrix T_1 and the operator $\mathcal{T}_2 : \mathbb{C}^{2n} \mapsto X_r$ by the partitioning*

$$\begin{bmatrix} T_1 \\ \mathcal{T}_2 \end{bmatrix} := \begin{bmatrix} \phi_{cl,1} & \phi_{cl,2} & \cdots & \phi_{cl,2n} \end{bmatrix} \tag{B.2}$$

and compute

$$\begin{bmatrix}\psi_1 & \psi_2 & \cdots & \psi_{2n}\end{bmatrix} := \begin{bmatrix} T_1^{-1} \\ -\mathcal{T}_2 T_1^{-1}\end{bmatrix}, \quad \psi_i \in X_{cl},\ i = 1, 2, \ldots, 2n. \tag{B.3}$$

Therein, T_1^{-1} can be calculated easily because T_1 is an invertible $2n \times 2n$-matrix.

3. *Calculate the $m \times 2n$-matrix $G = \mathcal{C}_r \mathcal{T}_2$ and define the vectors*

$$c_{e,i} := \begin{bmatrix} g_i \\ c_i \end{bmatrix}, \quad i = 1, 2, \ldots, m \tag{B.4}$$

with $c_1, c_2, \ldots, c_m$ from (2.15) and $g_1, g_2, \ldots, g_m$ from

$$\begin{bmatrix} g_1^T \\ \vdots \\ g_m^T \end{bmatrix} = \overline{G}. \tag{B.5}$$

4. *Compute the* Gramian matrices

$$\Pi := \begin{bmatrix} \langle \psi_1, \psi_1 \rangle_{X_{cl}} & \cdots & \langle \psi_n, \psi_1 \rangle_{X_{cl}} \\ \vdots & & \vdots \\ \langle \psi_1, \psi_n \rangle_{X_{cl}} & \cdots & \langle \psi_n, \psi_n \rangle_{X_{cl}} \end{bmatrix} \tag{B.6}$$

$$\Gamma := \begin{bmatrix} \langle c_{e,1}, c_{e,1} \rangle_{X_{cl}} & \cdots & \langle c_{e,m}, c_{e,1} \rangle_{X_{cl}} \\ \vdots & & \vdots \\ \langle c_{e,1}, c_{e,m} \rangle_{X_{cl}} & \cdots & \langle c_{e,m}, c_{e,m} \rangle_{X_{cl}} \end{bmatrix} \tag{B.7}$$

with the vectors in (B.3) and (B.4).

Then, the radius r in (2.204)–(2.205) of the enclosing circles is given by

$$r = \| \Pi^{\frac{1}{2}} L \Gamma^{\frac{1}{2}} \|_{\mathbb{C}^{n \times m}}, \tag{B.8}$$

where L is the observer gain of Σ_c (see (2.168)).

Note, that the boundary value problems in the first step of the procedure can be solved numerically because the assigned eigenvalues $\lambda_{cl,i} \in \sigma(A_n - B_n K) \cup \sigma(A_n - L C_n)$ are known (see (2.179)). The Steps 3 and 4 require the evaluation of several inner products on X_{cl} which involves in most cases a numerical integration. Thus, the presented algorithm can be implemented by numerical software packages.

In order to prove the statement Theorem 2.4-10 will be applied, in which the normalizing transformation $\mathcal{T}_{cl}$ can be shown to have a certain structure under the condition that $\mathcal{A}$ is normal. This is subject of the following auxiliary statement.

Lemma B-2

Assume that $\mathcal{A}$ is normal. Let T_1 and $\mathcal{T}_2$ be as defined in Theorem B-1. Then, the operator $\mathcal{T}_{cl} : X_{cl} \mapsto X_{cl}$ defined by

$$\mathcal{T}_{cl} \begin{bmatrix} h_{2n} \\ h_r \end{bmatrix} = \begin{bmatrix} T_1 & 0 \\ \mathcal{T}_2 & I_{X_r} \end{bmatrix} \begin{bmatrix} h_{2n} \\ h_r \end{bmatrix}, \qquad \forall\, h_{2n} \in \mathbb{C}^{2n},\, h_r \in X_r, \tag{B.9}$$

with I_{X_r} denoting the identity operator on X_r, is a linear transformation, and $\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl}$ is normal. Furthermore, $\mathcal{T}_{cl}^{-1}$ is given by

$$\mathcal{T}_{cl}^{-1} \begin{bmatrix} h_{2n} \\ h_r \end{bmatrix} = \begin{bmatrix} T_1^{-1} & 0 \\ -\mathcal{T}_2 T_1^{-1} & I_{X_r} \end{bmatrix} \begin{bmatrix} h_{2n} \\ h_r \end{bmatrix}, \qquad \forall\, h_{2n} \in \mathbb{C}^{2n},\, h_r \in X_r. \tag{B.10}$$

Proof. First, it can be verified easily under use of (B.9) that (B.10) holds. Since T_1, $\mathcal{T}_2$, and I_{X_r} are bounded operators, also $\mathcal{T}_{cl}$ and $\mathcal{T}_{cl}^{-1}$ are bounded so that $\mathcal{T}_{cl}$ is a linear transformation as claimed. Using the fact that the eigenvectors $\phi_{cl,i}$ satisfy the eigenvalue-eigenvector equation

$$\mathcal{A}_{cl,i}\phi_{cl,i} = \lambda_{cl,i}\phi_{cl,i}, \qquad i = 1, 2, \ldots, 2n, \tag{B.11}$$

one obtains from (B.2)

$$\mathcal{A}_{cl,0} \begin{bmatrix} T_1 \\ \mathcal{T}_2 \end{bmatrix} = \begin{bmatrix} T_1 \\ \mathcal{T}_2 \end{bmatrix} \Lambda_{2n} \tag{B.12}$$

with

$$\Lambda_{2n} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{2n}), \tag{B.13}$$

and thus

$$\mathcal{A}_{cl,0}\mathcal{T}_{cl} = \mathcal{A}_{cl,0} \begin{bmatrix} T_1 & 0 \\ \mathcal{T}_2 & I_{X_r} \end{bmatrix} = \begin{bmatrix} T_1 & 0 \\ \mathcal{T}_2 & I_{X_r} \end{bmatrix} \begin{bmatrix} \Lambda_{2n} & 0 \\ 0 & \mathcal{A}_r \end{bmatrix}. \tag{B.14}$$

by aid of (2.177). Combined with (B.10) this yields

$$\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl} = \begin{bmatrix} T_1^{-1} & 0 \\ -\mathcal{T}_2 T_1^{-1} & I_{X_r} \end{bmatrix} \begin{bmatrix} T_1 & 0 \\ \mathcal{T}_2 & I_{X_r} \end{bmatrix} \begin{bmatrix} \Lambda_{2n} & 0 \\ 0 & \mathcal{A}_r \end{bmatrix} = \begin{bmatrix} \Lambda_{2n} & 0 \\ 0 & \mathcal{A}_r \end{bmatrix} \tag{B.15}$$

and

$$(\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})^* = \begin{bmatrix} \Lambda_{2n}^* & 0 \\ 0 & \mathcal{A}_r^* \end{bmatrix}. \tag{B.16}$$

Using the fact that $\mathcal{A}_r$ is normal, *i.e.*, $\mathcal{A}_r\mathcal{A}_r^* = \mathcal{A}_r^*\mathcal{A}_r$, since $\mathcal{A}$ is normal by assumption and $\mathcal{A}_r h = \mathcal{A}h$, $\forall\, h \in D(\mathcal{A}_r)$ (see (2.115)), it follows from (B.15)–(B.16) and (B.13)

$$\begin{aligned}(\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})(\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})^* &= \begin{bmatrix}\Lambda_{2n}\Lambda_{2n}^* & 0\\ 0 & \mathcal{A}_r\mathcal{A}_r^*\end{bmatrix} = \begin{bmatrix}\Lambda_{2n}^*\Lambda_{2n} & 0\\ 0 & \mathcal{A}_r^*\mathcal{A}_r\end{bmatrix}\\ &= (\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})^*(\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl}) \qquad \text{(B.17)}\\ D((\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})(\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})^*) &= \mathbb{C}^{2n} \oplus D(\mathcal{A}_r\mathcal{A}_r^*) = \mathbb{C}^{2n} \oplus D(\mathcal{A}_r^*\mathcal{A}_r)\\ &= D((\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})^*(\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl})). \qquad \text{(B.18)}\end{aligned}$$

Thus, $\mathcal{T}_{cl}^{-1}\mathcal{A}_{cl,0}\mathcal{T}_{cl}$ is normal, so that the proof is complete. ■

Now, the proof of Theorem B-1 follows next.

Proof of Theorem B-1. In order to show (B.8) Theorem 2.4-10 is applied, for which the transformation $\mathcal{T}_{cl}$ according to (B.9) will be used. In order to avoid confusion the identity operator on X_r is denoted I_{X_r} in this proof. With the structure of Δ (see (2.177)) $\tilde{\Delta} \coloneqq \mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}$ can easily be verified to be given by consequently

$$\tilde{\Delta} = \begin{bmatrix}T_1^{-1}\\ -\mathcal{T}_2T_1^{-1}\end{bmatrix}\begin{bmatrix}-L\\ 0_{n\times m}\end{bmatrix}\mathcal{C}_r\begin{bmatrix}\mathcal{T}_2 & I_{X_r}\end{bmatrix} \qquad \text{(B.19)}$$

with $0_{n\times m}$ being the $n \times m$ zero matrix. Defining

$$\mathcal{C}_e \coloneqq \mathcal{C}_r\begin{bmatrix}\mathcal{T}_2 & I_{X_r}\end{bmatrix} \qquad \text{(B.20)}$$

$$\Psi \coloneqq \begin{bmatrix}\psi_1 & \psi_2 & \cdots & \psi_n\end{bmatrix}, \qquad \text{(B.21)}$$

where ψ_i, $i = 1, 2, \ldots, n$, are as in (B.3), the relation (B.19) becomes

$$\tilde{\Delta} = -\Psi\, L\, \mathcal{C}_e \qquad \text{(B.22)}$$

by aid of (B.3). Next, a connection between $\mathcal{C}_e$ and Ψ on the one hand and Γ and Π in (B.6)–(B.7) on the other will be examined. As a first step it has to be be shown that the operator $\mathcal{C}_e$ can be represented as

$$\mathcal{C}_e\begin{bmatrix}h_{2n}\\ h_r\end{bmatrix} = \begin{bmatrix}\langle\begin{bmatrix}h_{2n}\\ h_r\end{bmatrix}, c_{e,1}\rangle_{X_{cl}}\\ \vdots\\ \langle\begin{bmatrix}h_{2n}\\ h_r\end{bmatrix}, c_{e,m}\rangle_{X_{cl}}\end{bmatrix} = \begin{bmatrix}\langle h_{2n}, g_1\rangle_{\mathbb{C}^{2n}} + \langle h_r, c_1\rangle_{X_r}\\ \vdots\\ \langle h_{2n}, g_m\rangle_{\mathbb{C}^{2n}} + \langle h_r, c_m\rangle_{X_r}\end{bmatrix}, \qquad \forall\begin{bmatrix}h_{2n}\\ h_r\end{bmatrix} \in X_{cl} \qquad \text{(B.23)}$$

with $c_{e,1}, \ldots, c_{e,m}$ defined by (B.4)–(B.5). In order to prove this relation apply $\mathcal{C}_e$ to $\left[\begin{smallmatrix} h_{2n} \\ h_r \end{smallmatrix}\right] \in X_{cl}$ with $h_{2n} \in \mathbb{C}^{2n}$ and $h_r \in X_r$ yielding

$$\mathcal{C}_e \begin{bmatrix} h_{2n} \\ h_r \end{bmatrix} = \mathcal{C}_r \begin{bmatrix} \mathcal{T}_2 & I_{X_r} \end{bmatrix} \begin{bmatrix} h_{2n} \\ h_r \end{bmatrix} = G h_{2n} + \mathcal{C}_r h_r \tag{B.24}$$

(see (B.20)), where G is the $m \times 2n$-matrix defined in Step 3 of Theorem B-1. Using (B.5) leads to

$$G h_{2n} = \begin{bmatrix} \overline{g_1^T}\, h_{2n} \\ \vdots \\ \overline{g_m^T}\, h_{2n} \end{bmatrix} = \begin{bmatrix} \langle h_{2n}, g_1 \rangle_{\mathbb{C}^{2n}} \\ \vdots \\ \langle h_{2n}, g_m \rangle_{\mathbb{C}^{2n}} \end{bmatrix}, \tag{B.25}$$

and from (2.15) and (2.115) one gets

$$\mathcal{C}_r h_r = \begin{bmatrix} \langle h_r, c_1 \rangle_{X_r} \\ \vdots \\ \langle h_r, c_m \rangle_{X_r} \end{bmatrix}. \tag{B.26}$$

Then, insertion of (B.25)–(B.26) into (B.24) yields (B.23). It is straightforward to check by aid of (B.21) and (B.23) that the adjoint operators of Ψ and $\mathcal{C}_e$ are given by

$$\Psi^* h_r = \begin{bmatrix} \langle h_r, \psi_1 \rangle_{X_{cl}} \\ \vdots \\ \langle h_r, \psi_n \rangle_{X_{cl}} \end{bmatrix}, \qquad \forall\, h_r \in X_r \tag{B.27}$$

$$\mathcal{C}_e^* h_m = \begin{bmatrix} c_{e,1} & \cdots & c_{e,m} \end{bmatrix} h_m, \qquad \forall\, h_m \in \mathbb{C}^m, \tag{B.28}$$

and consequently the matrices Π and Γ in (B.6)–(B.7) satisfy

$$\Pi = \Psi^* \Psi, \qquad \Gamma = \mathcal{C}_e \mathcal{C}_e^*. \tag{B.29}$$

Note, that ψ_i, $i = 1, 2, \ldots, n$, are linearly independent because $\mathcal{T}_{cl}^{-1}$ is invertible (see (B.3) and (B.10)) so that any $\alpha \in \mathbb{C}^n \setminus \{0\}$ yields $\Psi\alpha \neq 0$. Similar, $c_{e,i}$, $i = 1, 2, \ldots, m$, are linearly independent as a consequence of $c_i, i = 1, 2, \ldots, m$, being linear independent (see Assumption 2.1-2), and this assures $\mathcal{C}_e^* \beta \neq 0$ for all $\beta \in \mathbb{C}^m \setminus \{0\}$. Using

$$\overline{\alpha^T} \Psi^* \Psi \alpha = (\Psi\alpha)^*\, \Psi\alpha = \langle (\Psi\alpha)^*\, \Psi\alpha, 1 \rangle_{\mathbb{C}} = \langle \Psi\alpha, \Psi\alpha \rangle_{X_{cl}} \tag{B.30}$$

$$\overline{\beta^T} \mathcal{C}_e \mathcal{C}_e^* \beta = (\mathcal{C}_e^* \beta)^*\, \mathcal{C}_e^* \beta = \langle (\mathcal{C}_e^* \beta)^*\, \mathcal{C}_e^* \beta, 1 \rangle_{\mathbb{C}} = \langle \mathcal{C}_e^* \beta, \mathcal{C}_e^* \beta \rangle_{X_{cl}} \tag{B.31}$$

the relations in (B.29) lead to

$$\overline{\alpha^T}\, \Pi\, \alpha = \overline{\alpha^T} \Psi^* \Psi \alpha = \langle \Psi\alpha, \Psi\alpha \rangle_{X_{cl}} = \|\Psi\alpha\|_{X_{cl}}^2 > 0 \qquad \forall\, \alpha \neq 0 \tag{B.32}$$

$$\overline{\beta^T}\, \Gamma\, \beta = \overline{\beta^T} \mathcal{C}_e \mathcal{C}_e^* \beta = \langle \mathcal{C}_e^* \beta, \mathcal{C}_e^* \beta \rangle_{X_{cl}} = \|\mathcal{C}_e^* \beta\|_{X_{cl}}^2 > 0 \qquad \forall\, \beta \neq 0, \tag{B.33}$$

showing that Π and Γ are positive definite matrices. In addition, these matrices are hermitian which is obvious in view of (B.29). For that reason there uniquely exist the square root matrices $\Pi^{\frac{1}{2}}$ and $\Gamma^{\frac{1}{2}}$ which are invertible because the square roots themselves are positive definite (see [21, Sec. 8.5]).

In order to avoid the computation of $\|\tilde{\Delta}\| = \|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\|$ in the infinite-dimensional space X_{cl} it is the aim in the sequel to show that the operators

$$\mathcal{H} := \Pi^{-\frac{1}{2}}\Psi^* \quad \text{and} \quad \mathcal{J} := \mathcal{C}_e^*\Gamma^{-\frac{1}{2}} \tag{B.34}$$

are *isometric* so that $\mathcal{H}\tilde{\Delta}\mathcal{J}$ is an $n \times m$-matrix that satisfies

$$\|\tilde{\Delta}\| = \|\mathcal{H}\tilde{\Delta}\mathcal{J}\|. \tag{B.35}$$

This allows to determine the norm of $\tilde{\Delta}$ easily as the norm of the $n \times m$-matrix $\mathcal{H}\tilde{\Delta}\mathcal{J}$. First, it will be verified that $\|\mathcal{H}\tilde{\Delta}\mathcal{J}\| = \|\tilde{\Delta}\mathcal{J}\|$ holds. In the light of (B.22), $\mathcal{H}\tilde{\Delta}\mathcal{J}\beta = -\mathcal{H}\Psi\, L\,\mathcal{C}_e\mathcal{J}\beta$ with $\beta \in \mathbb{C}^m$ can be written as

$$\mathcal{H}\tilde{\Delta}\mathcal{J}\beta = \mathcal{H}\Psi\alpha \tag{B.36}$$

with

$$\alpha = -L\mathcal{C}_e\mathcal{J}\beta \in \mathbb{C}^n, \tag{B.37}$$

and using (B.29) and (B.34) gives

$$\mathcal{H}\Psi\alpha = \Pi^{-\frac{1}{2}}\Pi\alpha = \Pi^{\frac{1}{2}}\alpha. \tag{B.38}$$

Consequently, by an analog reasoning as in (B.31), it follows with (B.36)

$$\|\mathcal{H}\tilde{\Delta}\mathcal{J}\beta\|_{\mathbb{C}^n}^2 = \|\mathcal{H}\Psi\alpha\|_{\mathbb{C}^n}^2 = \left\langle \Pi^{\frac{1}{2}}\alpha, \Pi^{\frac{1}{2}}\alpha \right\rangle_{\mathbb{C}^n} = \overline{\alpha^T}\left(\Pi^{\frac{1}{2}}\right)^*\Pi^{\frac{1}{2}}\alpha = \overline{\alpha^T}\,\Pi\,\alpha, \tag{B.39}$$

where the fact was used that $\Pi^{\frac{1}{2}}$ is hermitian since so is Π (see [21, Sec. 8.5]). Furthermore, in view of (B.22), (B.29), and (B.37), one obtains

$$\|\tilde{\Delta}\mathcal{J}\beta\|_{X_{cl}}^2 = \|-\Psi L\mathcal{C}_e\mathcal{J}\beta\|_{X_{cl}}^2 = \langle\Psi\alpha, \Psi\alpha\rangle_{X_{cl}} = \overline{\alpha^T}\Psi^*\,\Psi\alpha = \overline{\alpha^T}\,\Pi\,\alpha. \tag{B.40}$$

Since the relations (B.39)–(B.40) are valid of arbitrary $\beta \in \mathbb{C}^m$, comparison of (B.39) and (B.40) yields

$$\|\mathcal{H}\tilde{\Delta}\mathcal{J}\| = \|\tilde{\Delta}\mathcal{J}\|. \tag{B.41}$$

Now, it shall be shown that also $\mathcal{J}$ is isometric, *i.e.*, $\|\tilde{\Delta}\mathcal{J}\| = \|\tilde{\Delta}\|$. This can be done by an analog way as above, when it is used that $\|\tilde{\Delta}\mathcal{J}\| = \|(\tilde{\Delta}\mathcal{J})^*\|$ and $\|\tilde{\Delta}\| = \|\tilde{\Delta}^*\|$

holds in general (see [89, Sec. III 3.3]), so that it is sufficient to check the equality $\|\mathcal{J}^*\tilde{\Delta}^*\| = \|\tilde{\Delta}^*\|$. Making use of the defining relation (B.34) as well as (B.22) and (B.29) it holds

$$\mathcal{J}^*\tilde{\Delta}^* = -\left(\Gamma^{-\frac{1}{2}}\right)^* \mathcal{C}_e\mathcal{C}_e^* L^*\Psi^* = -\Gamma^{-\frac{1}{2}}\Gamma L^*\Psi^* = -\Gamma^{\frac{1}{2}} L^*\Psi^* \tag{B.42}$$

because $\Gamma^{-\frac{1}{2}}$ is hermitian. Consequently, for any $h \in X_{cl}$ it follows

$$\|\mathcal{J}^*\tilde{\Delta}^* h\|^2_{\mathbb{C}^m} = \left\langle -\Gamma^{\frac{1}{2}} L^*\Psi^* h, -\Gamma^{\frac{1}{2}} L^*\Psi^* h \right\rangle_{\mathbb{C}^m} = \overline{h^T}\Psi\, L\Gamma\, L^*\Psi^* h, \tag{B.43}$$

and comparison with

$$\|\tilde{\Delta}^* h\|^2_{X_{cl}} = \langle -\mathcal{C}_e^* L^*\Psi^* h, -\mathcal{C}_e^* L^*\Psi^* h\rangle_{X_{cl}} = \overline{h^T}\Psi L\mathcal{C}_e\mathcal{C}_e^* L^*\Psi^* h = \overline{h^T}\Psi\, L\Gamma\, L^*\Psi^* h \tag{B.44}$$

(see (B.22)) makes apparent that $\|\mathcal{J}^*\tilde{\Delta}^*\| = \|\tilde{\Delta}^*\|$. Thus, $\mathcal{J}$ has the desired property

$$\|\tilde{\Delta}\mathcal{J}\| = \|\tilde{\Delta}\|. \tag{B.45}$$

Hence, (B.41) and (B.45) give (B.35). By aid of (B.22), (B.29), and (B.34) it follows

$$\|\mathcal{H}\tilde{\Delta}\mathcal{J}\| = \| - \Pi^{-\frac{1}{2}}\Psi^*\Psi L\mathcal{C}_e\mathcal{C}_e^*\Gamma^{-\frac{1}{2}}\| = \|\Pi^{-\frac{1}{2}}\Pi L\Gamma\Gamma^{-\frac{1}{2}}\| = \|\Pi^{\frac{1}{2}} L\Gamma^{\frac{1}{2}}\|. \tag{B.46}$$

In view of $r = \|\mathcal{T}_{cl}^{-1}\Delta\mathcal{T}_{cl}\| = \|\tilde{\Delta}\|$ (see (2.205)), insertion of (B.46) in (B.35) finally completes the proof. ■

Appendix C

The adjoint operator

In this section it is demonstrated on the basis of an example how the adjoint $\mathcal{A}^*$ of a system operator $\mathcal{A}$ can be determined in many cases. For that purpose a system is considered that consists of two connected heat conducting rods as shown in Figure 39. The temperature distributions $\vartheta(z_i, t)$ along the heat conductors with spatial coordinates $z_1 \in [0, \ell_1]$ and $z_2 \in [0, \ell_2]$ are described by the homogeneous heat equations

$$\partial_t \vartheta_i(z_i, t) = \partial_{z_i}^2 \vartheta_i(z_i, t), \qquad t > 0,\ i = 1, 2,\ z_i \in (0, \ell_i) \tag{C.1}$$

(see Example 2.1-1). Since the heat flow

$$\Phi_i(z_i, t) = \partial_{z_i} \vartheta_i(z_i, t), \qquad i = 1, 2 \tag{C.2}$$

is considered to vanish at $z_1 = 0$ and $z_2 = \ell_2$ for all $t > 0$ *Neumann boundary conditions*

$$\partial_{z_1} \vartheta_1(0, t) = 0, \qquad t > 0 \tag{C.3}$$

$$\partial_{z_2} \vartheta_2(\ell_2, t) = 0, \qquad t > 0 \tag{C.4}$$

are used. In the following the shorthand notation $\vartheta_1'(z_1, t) := \partial_{z_1} \vartheta_1(z_1, t)$, $\vartheta_2'(z_2, t) := \partial_{z_2} \vartheta_2(z_2, t)$ is applied. The temperatures of both heat conductors must coincide at the connection point, and the energy flow $\Phi_1(\ell_1, t)$ out of heat conductor 1 at its right end has to equal the energy flow $\Phi_2(0, t)$ into the heat conductor 2 at the left end, which is described by

$$\vartheta_1(\ell_1, t) = \vartheta_2(0, t), \qquad t > 0 \tag{C.5}$$

$$\vartheta_1'(\ell_1, t) = \vartheta_2'(0, t), \qquad t > 0. \tag{C.6}$$

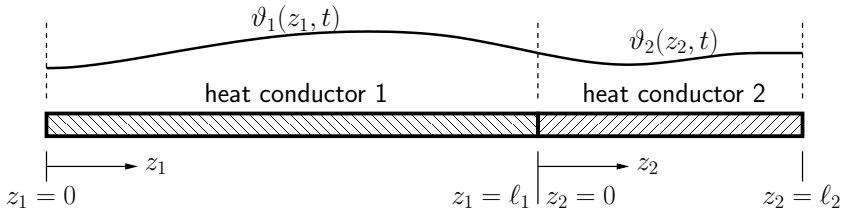

Figure 39 – Two connected heat conductors with lengths ℓ_1 and ℓ_2, and temperature distributions $\vartheta_1(z_1,t)$ and $\vartheta_2(z_2,t)$, respectively.

In conjunction with the initial condition $\vartheta_i(z_i,0) = \vartheta_{i,0}(z_i)$, $z_i \in [0,\ell_1]$, $i = 1,2$, (C.1)–(C.6) shall be described by a state space model with the state

$$x(t) = \begin{bmatrix} \vartheta_1(\cdot,t) \\ \vartheta_2(\cdot,t) \end{bmatrix} \tag{C.7}$$

on the state space $X = L_2(0,\ell_1) \oplus L_2(0,\ell_2)$, which is a Hilbert space with the inner product

$$\left\langle \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \right\rangle_X := \int_0^{\ell_1} h_1(z_1)\overline{g_1(z_1)}\mathrm{d}z_1 + \int_0^{\ell_2} h_2(z_2)\overline{g_2(z_2)}\mathrm{d}z_2. \tag{C.8}$$

Then, using (C.1)–(C.6), the system can be described by the homogeneous state equation

$$\dot{x}(t) = \mathcal{A}x(t), \qquad t > 0,\ x(0) = x_0 \in X, \tag{C.9}$$

where $\mathcal{A}$ is given by

$$\mathcal{A}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} h_1'' \\ h_2'' \end{bmatrix}, \qquad \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \in D(\mathcal{A}) \tag{C.10}$$

$$\begin{aligned} D(\mathcal{A}) = \big\{ & h_1 \in H_2(0,\ell_1),\ h_2 \in H_2(0,\ell_2) \ \big| \\ & h_1'(0) = 0, \\ & h_2'(\ell_2) = 0, \\ & h_1(\ell_1) = h_2(0), \\ & h_1'(\ell_1) = h_2'(0) \big\} \end{aligned} \tag{C.11}$$

which is the generator of a C_0-semigroup (see Appendix D for the definition of $H_2(0,\ell_i)$). Now, an *algebraic adjoint* of $\mathcal{A}$ shall be determined, which by definition is an operator

$$\tilde{\mathcal{A}}^* : D(\tilde{\mathcal{A}}^*) \subset X \mapsto X \tag{C.12}$$

such that

$$\langle \mathcal{A}h, g\rangle_X = \langle h, \tilde{\mathcal{A}}^* g\rangle_X, \qquad \forall\, h \in D(\mathcal{A}),\ g \in D(\tilde{\mathcal{A}}^*) \tag{C.13}$$

holds. Introducing $g = \left[\begin{smallmatrix} g_1 \\ g_2 \end{smallmatrix}\right]$, $h = \left[\begin{smallmatrix} h_1 \\ h_2 \end{smallmatrix}\right]$, and

$$\begin{bmatrix} \tilde{g}_1 \\ \tilde{g}_2 \end{bmatrix} = \tilde{g} = \tilde{\mathcal{A}}^* g \tag{C.14}$$

this means in view of (C.8) and (C.10) that $\tilde{g}_1$, $\tilde{g}_2$ must satisfy

$$\begin{aligned} \int_0^{\ell_1} h_1''(z_1)\overline{g_1(z_1)}\mathrm{d}z_1 + \int_0^{\ell_2} h_2''(z_2)\overline{g_2(z_2)}\mathrm{d}z_2 \\ = \int_0^{\ell_1} h_1(z_1)\overline{\tilde{g}_1(z_1)}\mathrm{d}z_1 + \int_0^{\ell_2} h_2(z_2)\overline{\tilde{g}_2(z_2)}\mathrm{d}z_2. \end{aligned} \tag{C.15}$$

Integration by parts, applied two times to the left-hand side of (C.15), yields

$$\begin{aligned} \int_0^{\ell_1} h_1''\overline{g_1}\mathrm{d}z_1 + \int_0^{\ell_2} h_2''\overline{g_2}\mathrm{d}z_2 = h_1'\overline{g_1}\big|_0^{\ell_1} - h_1\overline{g_1'}\big|_0^{\ell_1} + \int_0^{\ell_1} h_1\overline{g_1''}\mathrm{d}z_1 \\ + h_2'\overline{g_2}\big|_0^{\ell_2} - h_2\overline{g_2'}\big|_0^{\ell_2} + \int_0^{\ell_2} h_2\overline{g_2''}\mathrm{d}z_2, \end{aligned} \tag{C.16}$$

wherein the spatial arguments have been omitted for the sake of a simple notation. Relation (C.15) can be satisfied with $\tilde{g}_1, \tilde{g}_2 \in X$ only if the part

$$\begin{aligned} &h_1'\overline{g_1}\big|_0^{\ell_1} - h_1\overline{g_1'}\big|_0^{\ell_1} + h_2'\overline{g_2}\big|_0^{\ell_2} - h_2\overline{g_2'}\big|_0^{\ell_2} \\ &\quad = h_1'(\ell_1)\,\overline{g_1(\ell_1)} - h_1'(0)\,\overline{g_1(0)} - h_1(\ell_1)\,\overline{g_1'(\ell_1)} + h_1(0)\,\overline{g_1'(0)} \\ &\qquad + h_2'(\ell_2)\,\overline{g_2(\ell_2)} - h_2'(0)\,\overline{g_2(0)} - h_2(\ell_2)\,\overline{g_2'(\ell_2)} + h_2(0)\,\overline{g_2'(0)} \end{aligned} \tag{C.17}$$

in (C.16) vanishes. Inserting the boundary conditions (see (C.11)) this means that

$$h_1(0)\,\overline{g_1'(0)} - h_2(\ell_2)\,\overline{g_2'(\ell_2)} + h_2'(0)\big(\overline{g_1(\ell_1)} - \overline{g_2(0)}\big) + h_2(0)\big(\overline{g_2'(0)} - \overline{g_1'(\ell_1)}\big) \stackrel{!}{=} 0 \tag{C.18}$$

has to hold which can be satisfied for arbitrary $h_1(0)$, $h_2(\ell_2)$, $h_2'(0)$, and $h_2(0)$ only if the *adjoint boundary conditions*

$$g_1'(0) = 0 \tag{C.19}$$

$$g_2'(\ell_2) = 0 \tag{C.20}$$

$$g_1(\ell_1) = g_2(0) \tag{C.21}$$

$$g_1'(\ell_1) = g_2'(0) \tag{C.22}$$

hold. Under these conditions (C.16) simplifies to

$$\int_0^{\ell_1} h_1''\overline{g_1}\mathrm{d}z_1 + \int_0^{\ell_2} h_2''\overline{g_2}\mathrm{d}z_2 = \int_0^{\ell_1} h_1\overline{g_1''}\mathrm{d}z_1 + \int_0^{\ell_2} h_2\overline{g_2''}\mathrm{d}z_2 \tag{C.23}$$

which leads to

$$\tilde{\mathcal{A}}^* \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} g_1'' \\ g_2'' \end{bmatrix}, \qquad \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \in D(\tilde{\mathcal{A}}^*) \tag{C.24}$$

in view of (C.14)–(C.15). Before the domain of $\tilde{\mathcal{A}}^*$ is discussed it should be remarked that (C.24) defines an algebraic adjoint for any $D(\tilde{\mathcal{A}}^*)$ that incorporates the conditions (C.19)–(C.22) and that assures that the derivatives in (C.24) exist, since the defining relation (C.13) is then satisfied. Thus, an algebraic adjoint is not unique since any subset of this domain $D(\tilde{\mathcal{A}}^*)$ leads to another algebraic adjoint. In case however that $D(\tilde{\mathcal{A}}^*)$ is *maximal*[1], this special algebraic adjoint is called *adjoint*, which is denoted $\mathcal{A}^*$. Such an operator can be found if $\mathcal{A}$ is *densely defined*[2] (see [89, Sec. III 5.5]). Since g_1 and g_2 have to be sufficiently smooth for g_1'' and g_2'' in (C.24) to be defined, one could assume $g_1 \in C^2(0,\ell_1)$, $g_2 \in C^2(0,\ell_2)$ (see Appendix D for the definition of $C^2(0,\ell_2)$) which yields

$$\begin{aligned} D(\tilde{\mathcal{A}}^*) = \big\{ & g_1 \in C^2(0,\ell_1),\, g_2 \in C^2(0,\ell_2) \;\big|\; \\ & g_1'(0) = g_2'(\ell_2) = 0,\, g_1(\ell_1) = g_2(0),\, g_1'(\ell_1) = g_2'(0) \big\} \end{aligned} \tag{C.25}$$

in view of (C.19)–(C.22). However, the operator (C.24)–(C.25) can be extended because $g_1 \in H_2(0,\ell_1)$, $g_2 \in H_2(0,\ell_2)$ still yields $\tilde{\mathcal{A}}^* \left[\begin{smallmatrix} g_1 \\ g_2 \end{smallmatrix}\right] \in X$ as postulated in (C.12) so that a larger domain for $\tilde{\mathcal{A}}^*$ is achieved in view of $H_2(0,\ell_i) \supset C^2(0,\ell_i)$. In fact, it can be shown that the adjoint $\mathcal{A}^*$ is given by (C.24) and

$$\begin{aligned} D(\mathcal{A}^*) = \big\{ & g_1 \in H_2(0,\ell_1),\, g_2 \in H_2(0,\ell_2) \;\big|\; \\ & g_1'(0) = g_2'(\ell_2) = 0,\, g_1(\ell_1) = g_2(0),\, g_1'(\ell_1) = g_2'(0) \big\}. \end{aligned} \tag{C.26}$$

Comparison of $\mathcal{A}^*$ and $\mathcal{A}$ (see (C.10)–(C.11)) shows that both operators coincide in this example, *i.e.*,

$$\mathcal{A}h = \mathcal{A}^* h, \qquad h \in D(\mathcal{A}) \tag{C.27}$$

$$D(\mathcal{A}) = D(\mathcal{A}^*). \tag{C.28}$$

An operator $\mathcal{A}$ with adjoint $\mathcal{A}^*$ satisfying these two conditions is called *self-adjoint*. If $\mathcal{A}$ and $\mathcal{A}^*$ satisfy (C.27), but instead of (C.28) one has $D(\mathcal{A}) \subseteq D(\mathcal{A}^*)$, then $\mathcal{A}$ is called *symmetric*.

---

[1] This means that if $\tilde{\mathcal{A}}_1^*$ and $\tilde{\mathcal{A}}_2^*$ are algebraic adjoints of $\mathcal{A}$, in which $D(\tilde{\mathcal{A}}_1^*)$ is maximal, then $D(\tilde{\mathcal{A}}_1^*) \subseteq D(\tilde{\mathcal{A}}_2^*)$ implies $D(\tilde{\mathcal{A}}_1^*) = D(\tilde{\mathcal{A}}_2^*)$ and thus $\tilde{\mathcal{A}}_1^* = \tilde{\mathcal{A}}_2^*$.

[2] An operator $\mathcal{M} : D(\mathcal{M}) \subset X \mapsto X$ is said to be *densely defined* if $D(\mathcal{M})$ is dense in X, *i.e.*, $\overline{D(\mathcal{M})} = X$.

Appendix D

Definitions of function spaces

$C^k(a,b)$: **classical functions space**
Space of complex functions $f : [a,b] \mapsto \mathbb{C}$ over the interval $[a,b]$ for that the derivative $f^{(k)}$, $k \in \mathbb{N}_0$, is continuous on (a,b) and has a continuous extention on $[a,b]$.

$C^\infty(a,b)$: **classical functions space**
Space of complex functions $f : [a,b] \mapsto \mathbb{C}$ over the interval $[a,b]$ for that *any* derivative $f^{(k)}$, $\forall\, k \in \mathbb{N}_0$, is continuous on (a,b) and has a continuous extention on $[a,b]$.

$L_2(\Omega)$: **Lebesgue space**
Space of complex functions $f : \Omega \mapsto \mathbb{C}$ over the compact set Ω that are absolute square Lebesgue integrable, *i.e.*,

$$\int_\Omega |f(\omega)|^2 \mathrm{d}\omega < \infty.$$

$L_2(\Omega)$ is a Hilbert space with the inner product

$$\langle g, h \rangle_{L_2} := \int_\Omega g(\omega)\overline{h(\omega)}\mathrm{d}\omega.$$

$L_2(a,b)$: **Lebesgue space**

$$L_2(a,b) := L_2(\Omega) \quad \text{with} \quad \Omega = [a,b]$$

$L_p([a,b];\mathbb{C}^n)$: **Lebesgue space**

Space of $\mathbb{C}^n$-valued functions $f : [a,b] \mapsto \mathbb{C}^n$ over the interval $[a,b] \subseteq \mathbb{R}$ that are absolute Lebesgue p-integrable with $p \in \mathbb{N}$, *i.e.*,

$$\int_a^b \|f(\omega)\|_{\mathbb{C}^n}^p \mathrm{d}\omega < \infty.$$

$L_2([a,b];\mathbb{C}^n)$ (*i.e.*, $p = 2$) is a Hilbert space with the inner product

$$\langle g, h \rangle_{L_2([a,b];\mathbb{C}^n)} := \int_a^b g^T(\omega)\overline{h(\omega)}\mathrm{d}\omega.$$

$L_\infty(a,b)$: **Lebesgue space**

Space of $\mathbb{C}$-valued functions $f : [a,b] \mapsto \mathbb{C}$ over the interval $[a,b] \subset \mathbb{R}$ that satisfy

$$|f(\omega)| < \infty, \qquad \forall\, \omega \in [a,b].$$

$W_{m,p}(\Omega)$: **Sobolov space**

Space of complex functions $f : \Omega \mapsto \mathbb{C}$ over the compact set Ω that and their first m derivatives are absolute Lebesgue p-integrable. To be more precise, it holds

$$\int_\Omega \sum_{i=0}^{m} |f^{(i)}(z)|^p \mathrm{d}\omega < \infty,$$

where the derivatives are to be taken in the weak sense.

$W_{m,p}(a,b)$: **Sobolev space**

$$W_{m,p}(a,b) := W_{m,p}(\Omega) \quad \text{with} \quad \Omega = [a,b]$$

$W_{m,p}([a,b];\mathbb{C}^n)$: **Sobolev space**

Space of complex functions $f : [a,b] \mapsto \mathbb{C}^n$ over the interval $[a,b] \subseteq \mathbb{R}$ that and their first m derivatives are absolute Lebesgue p-integrable. To be more precise, it holds

$$\int_a^b \sum_{i=0}^{m} \|f^{(i)}(\omega)\|_{\mathbb{C}^n}^p \mathrm{d}\omega < \infty,$$

where the derivatives are to be taken in the weak sense.

$H_2(\Omega)$: **Sobolev space**

$$H_2(\Omega) \coloneqq W_{2,2}(\Omega).$$

$H_2(\Omega)$ is a Hilbert space with the inner product

$$\langle g, h \rangle_{H_2} \coloneqq \int_\Omega \sum_{i=0}^{2} g^{(i)}(\omega)\overline{h^{(i)}(\omega)}\mathrm{d}\omega.$$

$H_2(a,b)$: **Sobolev space**

$$H_2(a,b) \coloneqq H_2(\Omega) \quad \text{with} \quad \Omega = [a,b]$$

l_2 : **Sequence space**

Space of complex-valued sequences that are absolute square summable, *i.e.*,

$$l_2 \coloneqq \left\{ (\nu_i)_{i\in\mathbb{N}} \,|\, \sum_{i=1}^{\infty} |\nu_i|^2 < \infty \right\}$$

l_2 is a Hilbert space with the inner product

$$\langle g, h \rangle_{l_2} \coloneqq \sum_{i=1}^{\infty} g_i \overline{h_i}.$$

Bibliography

[1] Antoulas A. C. *Approximation of Large-Scale Dynamical Systems*. SIAM, Philadelphia, USA, 2005.

[2] Balas M. J. *Active control of flexible systems*. Journal of Optimization Theory and Applications 25 (1978), 415–436.

[3] Balas M. J. *Feedback control of flexible systems*. IEEE Transactions on Automatic Control 23 (1978), 673–679.

[4] Balas M. J. *Modal control of certain flexible dynamic systems: Estimates of residual mode effects*. In: Proceedings of the IEEE Conference on Decision and Control (1978), 237–241.

[5] Balas M. J. *Toward a more practical control theory for distributed parameter systems*. In: *Control and Dynamic Systems*, C. T. Leondes (Ed.), vol. 18. Academic Press, New York, USA, 1982, 361–421.

[6] Balas M. J. *Trends in large space structure control theory: Fondest hopes, wildest dreams*. IEEE Transactions on Automatic Control 27 (1982), 522–535.

[7] Balas M. J. *The Galerkin method and feedback control of linear distributed parameter systems*. Journal of Mathematical Analysis and Applications 91 (1983), 527–546.

[8] Balas M. J. *The mathematical structure of the feedback control problem for linear distributed parameter systems with finite-dimensional controllers*. In: *Control Theory for Distributed Parameter Systems and Applications*, F. Kappel, K. Kunisch, and W. Schappacher (Eds.). Springer, Berlin, Germany, 1983.

[9] Balas M. J. *The structure of discrete-time finite-dimensional control of distributed parameter systems*. Journal of Mathematical Analysis and Applications 102 (1984), 519–538.

[10] Balas M. J. *Suboptimality and stability of linear distributed-parameter systems with finite-dimensional controllers*. Journal of Optimization Theory and Applications 45 (1985), 1–19.

[11] Balas M. J. *Exponentially stabilizing finite-dimensional controllers for linear distributed parameter systems: Galerkin approximation of infinite dimensional controllers*. Journal of Mathematical Analysis and Applications 117 (1986), 358–384.

[12] Balas M. J. *Finite-dimensional control of distributed parameter systems by Galerkin approximation of infinite dimensional controllers*. Journal of Mathematical Analysis and Applications 114 (1986), 17–36.

[13] Balas M. J. *Finite-dimensional controllers for linear distributed parameter systems: Exponential stability using residual mode filters*. Journal of Mathematical Analysis and Applications 133 (1988), 283–296.

[14] Balas M. J. *Stable feedback control of linear distributed parameter systems: Time and frequency domain conditions*. Journal of Mathematical Analysis and Applications 225 (1998), 144–167.

[15] Banks H. T. (Ed.). *Control and Estimation in Distributed Parameter Systems*. SIAM, Philadelphia, USA, 1992.

[16] Banks H. T. and Kunisch K. *Estimation Techniques for Distributed Parameter Systems*. Birkhäuser, Boston, USA, 1989.

[17] Banks H. T., Smith R. C. and Wang Y. *Smart Material Structures—Modeling, Estimation and Control*. John Wiley & Sons, New York, USA, 1996.

[18] Bauer F. and Fike C. *Norms and exclusion theorems*. Numerische Mathematik 2 (1960), 137–141.

[19] Becker J., Meurer T. and Gaul L. *Flatness-based feedforward control design for flexible structures*. In: Proceedings of the IEEE International Conference on Control Applications (2006), 650–655.

[20] Bensoussan A., Prato G. D., Delfour M. C. and Mitter S. K. *Representation and Control of Infinite Dimensional Systems*. Birkhäuser, Boston, USA, 2007.

[21] Bernstein D. *Matrix Mathematics*. Princeton University Press, Princeton,

USA, 2005.

[22] Bernstein D. S. and Hyland D. C. *The optimal projection equations for finite-dimensional fixed-order dynamic compensation of infinite-dimensional systems*. SIAM Journal on Control and Optimization 24 (1986), 122–151.

[23] Bontsema J. and Curtain R. F. *A note on spillover and robustness for flexible systems*. IEEE Transactions on Automatic Control 33 (1988), 567–569.

[24] Boyd S. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia, USA, 1994.

[25] Bradshaw A. *Modal control of distributed-parameter vibratory systems*. International Journal of Control 19 (1974), 957–968.

[26] Bradshaw A. and Porter B. *Modal control of a class of distributed-parameter systems: Multi-eigenvalue assignment*. International Journal of Control 16 (1972), 277–285.

[27] Byrnes C. I., Gilliam D. S. and Shubov V. I. *Example of output regulation for a system with unbounded inputs and outputs*. In: Proceedings of the Conference on Decision and Control (1999), 4280–4284.

[28] Chen G. and Russell D. L. *A mathematical model for linear elastic systems with structural damping*. Quarterly of Applied Mathematics 40 (1982), 433–454.

[29] Chen S., Liu K. and Liu Z. *Spectrum and stability for elastic systems with global or local Kelvin-Voigt damping*. SIAM Journal on Applied Mathematics 59 (1998), 651–668.

[30] Chen Y. *Vibrations: Theoretical Methods*. Addison-Wesley Publishing Company, London, UK, 1966.

[31] Choi J. and Park U. *Spillover suppression via eigenstructure assignment in large flexible structures*. Journal of Guidance, Control, and Dynamics 25 (2002), 599–602.

[32] Cimino M. and Pagilla P. R. *Design of linear time-invariant controllers for multirate systems*. Automatica 46 (2010), 1315–1319.

[33] Conway J. B. *Functions of One Complex Variable I*. Springer, New York, USA, 1978.

[34] Cooper D. J., Ramirez W. F. and Clough D. E. *Comparison of linear distributed-parameter filters to lumped approximants*. AIChE Journal 32 (1986), 186–194.

[35] Courant R. and Hilbert D. *Methods of mathematical physics, Volume I*. Interscience Publishers, New York, USA, 1965.

[36] Curtain R. F. *The spectrum determined growth assumption for perturbations of analytic semigroups*. IEEE Transactions on Automatic Control 2 (1982), 106–109.

[37] Curtain R. F. *Compensators for infinite dimensional linear systems*. Journal of the Franklin Institute 315 (1983), 331–346.

[38] Curtain R. F. *Finite dimensional compensators for parabolic distributed systems with unbounded control and observation*. SIAM Journal on Control and Optimization 22 (1984), 255–276.

[39] Curtain R. F. *Spectral systems*. International Journal of Control 39 (1984), 657–666.

[40] Curtain R. F. *Pole assignment for distributed systems by finite-dimensional control*. Automatica 21 (1985), 57–67.

[41] Curtain R. F. *A note on spillover and robustness for flexible systems*. IEEE Transactions on Automatic Control 33 (1988), 567–569.

[42] Curtain R. F. *A comparison of finite-dimensional controller designs for distributed parameter systems*. Control-Theory and Advanced Technology 9 (1993), 609–628.

[43] Curtain R. F. *Transfer functions of distributed parameter systems: A tutorial*. Automatica 45 (2009), 1101–1116.

[44] Curtain R. F. and Weiss G. *Well posedness of triples of operators*. In: *Control and Estimation of Distributed Parameter Systems*, F. Kappel, K. Kunisch, and W. Schappacher (Eds.). Birkhäuser, Basel, Switzerland, 1989, 41–59.

[45] Curtain R. F. and Zwart H. J. *Functional Analysis in Modern Applied Mathematics*. Academic Press, London, UK, 1977.

[46] Curtain R. F. and Zwart H. J. *An Introduction to Infinite-Dimensional*

Linear Systems Theory. Springer, New York, USA, 1995.

[47] Damerau J. *Untersuchung der dynamischen Eigenschaften von Balken mit fraktionalen Stoffgesetzen (in German)*. PhD thesis, Institute of Mechanical Engineering, Helmut-Schmidt University Hamburg, Germany, 2008.

[48] Davison E. J. *The robust control of a servomechanism problem for linear time-invariant multivariable systems*. IEEE Transactions on Automatic Control 21 (1976), 25–34.

[49] Delattre C., Dochain D. and Winkin J. *Sturm-Liouville systems are Riesz-spectral systems*. International Journal of Applied Mathematics and Computer Science 13 (2003), 481–484.

[50] Deutscher J. *Output regulation for linear distributed-parameter systems using finite-dimensional dual observers*. Automatica 47 (2011), 2468–2473.

[51] Deutscher J. *Zustandsregelung verteilt-parametrischer Systeme (in German)*. Springer, Berlin, Germany, 2012.

[52] Deutscher J. *Finite-dimensional dual state feedback control of linear boundary control systems*. International Journal of Control 86 (2013), 41–53.

[53] Deutscher J. and Harkort C. *Parametric state feedback design for second-order Riesz-spectral systems*. In: Proceedings of the 10th European Control Conference (2009), 282–287.

[54] Deutscher J. and Harkort C. *Parametric state feedback design of linear distributed-parameter systems*. International Journal of Control 82 (2009), 1060–1069.

[55] Deutscher J. and Harkort C. *Finite-dimensional control of linear distributed-parameter systems using output observation (in German)*. at-Automatisierungstechnik 58 (2010), 435–446.

[56] Deutscher J. and Harkort C. *A parametric approach to finite-dimensional control of linear distributed-parameter systems*. International Journal of Control 83 (2010), 1674–1685.

[57] Deutscher J. and Harkort C. *Parametric approach to the decoupling of linear distributed-parameter systems*. IET Control Theory and Applications 4 (2010), 2855–2866.

[58] Deutscher J. and Harkort C. *Reference and disturbance feedforward controllers for linear distributed-parameter systems (in German).* at-Automatisierungstechnik 58 (2010), 27–37.

[59] Dunford N. and Schwartz J. *Linear Operators.* John Wiley & Sons, New York, USA, 1976.

[60] Engel K.-J. and Nagel R. *One-Parameter Semigroups for Linear Evolution Equations.* Springer, New York, USA, 2000.

[61] Fadali M. S. and Visioli A. *Digital Control Engineering.* Academic Press, Waltham, USA, 2012.

[62] Fahmy M. M. and O'Reilly J. *On eigenstructure assignment in linear multivariable systems.* IEEE Transactions on Automatic Control 27 (1982), 690–693.

[63] Finlayson B. A. *The Method of Weighted Residuals and Variational Principles.* Academic Press, New York, USA, 1972.

[64] Fletcher C. A. J. *Computational Galerkin Methods.* Springer, New York, USA, 1984.

[65] Francis B. A. and Wonham W. M. *The internal model principle of control theory*. Automatica 12 (1976), 457–465.

[66] Franke D. *Stability analysis via eigenvalue enclosure in distributed parameter systems.* In: *Complex and Distributed Systems: Analysis, Simulation and Control*, S. G. Tzafestas and P. Borne (Eds.). IMACS, Amsterdam, Netherlands, 1986, 223–228.

[67] Franke D. *Systeme mit örtlich verteilten Parametern (in German).* Springer, Berlin, Germany, 1987.

[68] Franke D. *On the control of infinite-dimensional systems via finite-dimensional observers.* In: Proceedings of the IASTED International Symposium Modelling, Identification and Control (1991).

[69] Gohberg I. C. and Krein M. G. *Introduction to the Theory of Linear Nonselfadjoint Operators.* American Mathematical Society, Providence, USA, 1969.

[70] Guo B.-Z. *Riesz basis generation, eigenvalues distribution, and exponential stability for a Euler-Bernoulli beam with joint feedback control.* Revista matemática

complutense XIV (2001), 205–229.

[71] Guo B.-Z. *Basis property of a Rayleigh beam with boundary stabilization.* Journal of Optimization Theory and Applications 112 (2002), 529–547.

[72] Guo B.-Z. and Yu R. *The Riesz basis property of dicrete operators and application to a Euler-Bernoulli beam equation with boundary linear feedback control.* IMA Journal of Mathematical Control and Information 18 (2001), 241–251.

[73] Guo B.-Z. and Zwart H. J. *Riesz spectral systems.* Technical report, Faculty of Mathematical Sciences, University of Twente, The Netherlands, 2001.

[74] Hairer E., Hørsett S. P. and Wanner G. *Solving Ordinary Differential Equations I.* Springer, Berlin, Germany, 1993.

[75] Han S. M., Benaroya H. and Wei T. *Dynamics of transversely vibrating beams using four engineering theories.* Journal of Sound and Vibration 225 (1999), 935–988.

[76] Harkort C. and Deutscher J. *Reduktion von Spillover für lineare verteilt-parametrische Systeme (in German).* In: *Methoden und Anwendungen der Regelungstechnik, Erlangen-Münchener Workshops 2007 und 2008*, G. Roppenecker and B. Lohmann (Eds.). Shaker, Aachen, Germany, 2009, 15–26.

[77] Harkort C. and Deutscher J. *Spillover reduction for linear distributed-parameter systems using dynamic extensions.* In: Proceedings of the 10th European Control Conference (2009), 294–299.

[78] Harkort C. and Deutscher J. *Abtastregelung linearer verteilt-parametrischer Systeme unter Verwendung allgemeiner Halteglieder (in German).* In: *Methoden und Anwendungen der Regelungstechnik, Erlangen-Münchener Workshops 2009 und 2010*, G. Roppenecker and B. Lohmann (Eds.). Shaker, Aachen, Germany, 2011, 89–99.

[79] Harkort C. and Deutscher J. *Discrete-time modal state reconstruction for infinite-dimensional systems using generalized sampling.* In: Proceedings of the 18th IFAC World Congress (2011), 13311–13316.

[80] Harkort C. and Deutscher J. *Finite-dimensional observer-based control of linear distributed parameter systems using cascaded output observers.* Interna-

tional Journal of Control 84 (2011), 107–122.

[81] Harkort C. and Deutscher J. *Finite-dimensional observer-based control of Riesz-spectral systems — spectrum properties and eigenvalue estimates*. Technical report, Chair of Automatic Control, University Erlangen-Nuremberg, Germany, 2011.

[82] Harkort C. and Deutscher J. *Krylov subspace methods for linear infinite-dimensional systems*. IEEE Transactions on Automatic Control 56 (2011), 441–447.

[83] Hille E. and Phillips R. S. *Functional Analysis and Semigroups*. American Mathematical Society Colloquium Publications, Providence, USA, 1957.

[84] Ito K. *Finite-dimensional compensators for infinite-dimensional systems via Galerkin-type approximations*. SIAM Journal on Control and Optimization 28 (1990), 1251–1269.

[85] Junkins J. L. *Introduction to Dynamics and Control of Flexible Structures*. Education Series, Washington D.C., USA, 1993.

[86] Kailath T. *Linear Systems*. Prentice Hall, Upper Saddle River, USA, 1980.

[87] Kappel F., Kunisch K. and Schappacher W. (Eds.). *Control Theory for Distributed Parameter Systems and Applications*. Springer, Berlin, Germany, 1983.

[88] Kappel F., Kunisch K. and Schappacher W. (Eds.). *Control and Estimation of Distributed Parameter Systems*. Birkhäuser, Basel, Switzerland, 1989.

[89] Kato T. *Perturbation Theory for Linear Operators*. Springer, Berlin, Germany, 1995.

[90] Knopp K. *Theory of Functions*. Dover Publications, Mineola, USA, 1996.

[91] Kreith F. and Bohn M. *Principles of Heat Transfer*. Thomson, Toronto, Canada, 2001.

[92] Kreyszig E. *Introductory functional analysis with applications*. John Wiley & Sons, New York, USA, 1989.

[93] Kučera V. *Analysis and Design of Discrete Linear Control Systems*. Prentice Hall, Hemel Hempstead, UK, 1991.

[94] Leondes C. T. *Control and Dynamic Systems.* Academic Press, New York, USA, 1982.

[95] LeVeque R. J. *Finite Difference Methods for Ordinary and Partial Differential Equations.* SIAM, Philadelphia, USA, 2007.

[96] Lindner D. K., Babendreier J. and Hamdan A. M. A. *Measures of controllability and observability and residues.* IEEE Transactions on Automatic Control 34 (1989), 648–650.

[97] Logemann H. *Stability and stabilizability of linear infinite-dimensional discrete-time systems.* IMA Journal of Mathematical Control and Information 9 (1992), 255–263.

[98] Logemann H., Rebarber R. and Townley S. *Generalized sampled-data stabilization of well-posed linear infinite-dimensional systems.* SIAM Journal on Control and Optimization 44 (2005), 1345–1369.

[99] Luenberger D. G. *An introduction to observers.* IEEE Transactions on Automatic Control 16 (1971), 596–602.

[100] Lumer G. and Rosenblum M. *Linear operator equations.* Proceedings of the American Mathematical Society 10 (1959), 32–41.

[101] Luo Z.-H., Guo B.-Z. and Morgül O. *Stability and Stabilization of Infinite Dimensional Systems with Applications.* Springer, New York, USA, 1999.

[102] Mattheij R. R. M., Rienstra S. W. and Boonkkamp J. *Partial Differential Equations — Modeling, Analysis, Computation.* SIAM, Philadelphia, USA, 2005.

[103] Meirovitch L. *Analytical Methods in Vibrations.* The MacMillan Company, New York, USA, 1967.

[104] Meirovitch L. *Dynamics and Control of Structures.* John Wiley & Sons, New York, USA, 1990.

[105] Meirovitch L. and Baruh H. *On the problem of observation spillover in self-adjoint distributed-parameter systems.* Journal of Optimization Theory and Applications 39 (1983), 269–291.

[106] Naylor A. W. and Sell G. R. *Linear Operator Theory in Engineering and*

Science. Springer, New York, USA, 1982.

[107] Pazy A. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, New York, USA, 1983.

[108] Porter B. and Bradshaw A. *Modal control of a class of distributed-parameter systems*. International Journal of Control 15 (1972), 673–681.

[109] Preumont A. *Vibration Control of Active Structures - An Introduction*. Kluwer Academic Publishers, Dordrecht, Netherlands, 2002.

[110] Rebarber R. L. *Spectral determination for a cantilever beam*. IEEE Transactions on Automatic Control 34 (1989), 502–510.

[111] Rebarber R. L. and Townley S. *Generalized sampled data feedback control of distributed parameter systems*. Systems and Control Letters 34 (1998), 229–240.

[112] Reddy B. D. *Introductory Functional Analysis*. Springer, New York, USA, 1998.

[113] Reisenauer B. T., Balas M. J. and Ramey M. *Reduced-order model based control of large flexible manipulators: Theory and experiments*. In: Proceedings of the American Control Conference (1990), 1760–1765.

[114] Roppenecker G. *State feedback control of linear systems — a renewed approach (in German)*. at – Automatisierungstechnik 57 (2009), 491–498.

[115] Rosen I. G. and Wang C. *On stabilizability and sampling for infinite dimensional systems*. IEEE Transactions on Automatic Control 37 (1992), 1653–1656.

[116] Rudin W. *Principles of Mathematical Analysis*. McGraw-Hill, New York, USA, 1976.

[117] Rudin W. *Principles of Mathematical Analysis*. Oldenbourg, München, Germany, 2009.

[118] Rudolph J. and Woittennek F. *Motion planning and open loop control design for linear distributed parameter systems with lumped controls*. International Journal of Control 81 (2008), 457–474.

[119] Salt J. and Albertos P. *Model-based multirate controllers design*. IEEE Transactions on Control Systems Technology 13 (2005), 988–997.

[120] Schumacher J. M. *Dynamic Feeedback in Finite- and Infinite-Dimensional Systems*. Mathematical Center, Amsterdam, Netherlands, 1981.

[121] Schumacher J. M. *A direct approach to compensator design for distributed parameter systems*. SIAM Journal of Control and Optimization 21 (1983), 823–837.

[122] Schumacher J. M. *Finite-dimensional regulators for a class of infinite-dimensional systems*. Systems and Control Letters 3 (1983), 7–12.

[123] Staffans O. J. *Well-Posed Linear Systems*. Cambridge University Press, Cambridge, UK, 2005.

[124] Sun S.-H. *On spectrum distribution of completely controllable linear systems*. SIAM Journal of Control and Optimization 19 (1981), 730–743.

[125] Timoshenko S., Young D. H. and Weaver W. *Vibration Problems in Engineering*. John Wiley & Sons, New York, USA, 1974.

[126] Triggiani R. *On the stabilizability problem in Banach spaces*. Journal of Mathematical Analysis and Applications 52 (1975), 383–403.

[127] Tucsnak M. and Weiss G. *Observation and Control for Operator Semigroups*. Birkhäuser, Boston, USA, 2009.

[128] Villaggio P. *Mathematical Models for Elastic Structures*. Cambridge University Press, Cambridge, USA, 1997.

[129] Wang P. K. C. *Modal feedback stabilization of a linear distributed system*. IEEE Transactions on Automatic Control 17 (1972), 552–553.

[130] Weiss G. *The representation of regular linear systems on Hilbert spaces*. In: *Control and Estimation of Distributed Parameter Systems*, F. Kappel, K. Kunisch, and W. Schappacher (Eds.). Birkhäuser, Basel, Switzerland, 1989, 401–416.

[131] Xu C.-Z. *On spectrum and Riesz basis assignment of infinite-dimensional linear systems by bounded linear feedbacks*. SIAM Journal on Control and Optimization 34 (1996), 521–541.

[132] Xu G.-Q. and Feng D.-X. *On the spectrum determined growth assumption and the pertarbation of C_0 semigroups*. Integral Equations and Operator Theory

39 (2001), 363–376.

[133] Zabczyk J. *On decomposition of generators*. SIAM Journal of Control and Optimization 16 (1978), 523–434.

[134] Zeidler E. *Introduction to Applied Functional Analysis*. Springer, New York, USA, 1995.

Index

Controllers for systems with spatially distributed parameters are often designed on the basis of an approximation model. Doing so, the neglected system dynamics might have a negative impact on the closed-loop behavior. This dissertation presents several approaches to deal with this undesired effect in a systematic way. These address continuous-time as well as discrete-time compensators with particular low order.

The considered system class covers a variety of linear parabolic, hyperbolic, and biharmonic systems with spatially distributed inputs and outputs, as well as a class of time-delay systems. The dynamics of these systems and the closed-loop behavior are represented on the basis of state space models. The applied functional analytic tools are explained in detail.

ISBN 978-3-944057-14-9

9 783944 057149

FAU UNIVERSITY PRESS 2014