Chair of Automatic Control

Technische Universität München

Philipp Seiwald

Adaptive Model Order Reduction of Structured DAEs by Krylov Subspace Methods

Master's thesis

30 September 2016

Supervisors:

Alessandro Castagnotto, M.Sc.
Maria Cruz Varona, M.Sc.
Prof. Dr.-Ing. habil. Boris Lohmann

Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbstständig durchgeführt zu haben und keine weiteren Hilfsmittel und Quellen als die angegebenen genutzt zu haben. Mit ihrer unbefristeten Aufbewahrung in der Lehrstuhlbibliothek erkläre ich mich einverstanden.

Garching bei München, den 30. September 2016 ____________________ (Philipp Seiwald)

Chair of Automatic Control (Prof. Dr.-Ing. habil. Boris Lohmann)
Technische Universität München
Boltzmannstraße 15
85748 Garching bei München
Germany

Lehrstuhl für Regelungstechnik (Prof. Dr.-Ing. habil. Boris Lohmann)
Technische Universität München
Boltzmannstraße 15
85748 Garching bei München
Deutschland

Abstract

In this work model order reduction (MOR) of differential algebraic equations (DAEs), with focus on structured, linear time-invariant systems, is investigated. As usual in the DAE-setting it is assumed, that the spectral projectors, which describe the structure of the model, are available. Those allow a separation of the actual dynamics and the involved algebraic equations which describe a constraint manifold to which the dynamics are bounded.

While $\mathcal{H}_2$ pseudo-optimal reduction by the Krylov-based pseudo-optimal rational Krylov (PORK) algorithm [42] is applied to the strictly proper part of the transfer function, Lyapunov balanced truncation (BT) according to [9] is used to find a minimal realization of the improper contribution. For this purpose the PORK algorithm, originally developed for the reduction of ordinary differential equations, is revisited in the context of strictly proper DAEs. As the original proof has to be modified, a detailed derivation of the PORK algorithm is presented. Furthermore the combination with adaptive MOR schemes like the stability-preserving, adaptive rational Krylov (SPARK) algorithm [30] and the cumulative reduction (CURE) framework [42] is discussed.

One of the main tools used in this thesis are generalized Sylvester equations. It is shown that they can be used to describe tangential-input rational Krylov subspaces, even in the case of a singular descriptor matrix. Moreover a formulation of the $\mathcal{H}_2$ inner-product of the transfer functions of two strictly proper DAEs via projected generalized Sylvester equations is presented. Those results are essential for the proof of $\mathcal{H}_2$ pseudo-optimality in PORK, but may also be useful in different contexts.

Finally an efficient overall-algorithm for the reduction of structured linear DAE-systems of arbitrary index is presented, which adaptively chooses appropriate interpolation data and reduced order. By means of several physically based models it is shown, that the proposed technique is applicable to common technical problems, regardless of properness or index of the given system. Furthermore the reduction of an artificially generated high-index system is demonstrated.

Kurzreferat

Diese Arbeit behandelt die Modellordnungsreduktion (MOR) von differential-algebraischen Gleichungssystemen (DAEs), wobei der Fokus auf strukturierten und linear-zeitinvarianten Systemen liegt. Wie üblich bei der Reduktion von DAEs wird davon ausgegangen, dass die spektralen Projektoren, welche die Struktur des Modells beschreiben, zur Verfügung stehen. Diese erlauben eine Aufteilung in die eigentliche Dynamik und einen Satz von algebraischen Gleichungen, welche eine Mannigfaltigkeit bestimmen auf die die Dynamik beschränkt ist.

Während auf den streng properen Anteil der komplexen Übertragungsfunktion $\mathcal{H}_2$ pseudo-optimale Reduktion durch den Krylow-basierten pseudo-optimalen rationalen Krylow (PORK) Algorithmus [42] angewendet wird, stellt balanciertes Abschneiden (BT) nach [9] eine Minimalrealisierung des nicht-properen Beitrags zur Verfügung. Hierzu wird der PORK Algorithmus, ursprünglich für gewöhnliche Differentialgleichungssysteme entwickelt, für die Anwendung auf streng propere DAE-Systeme untersucht. Da der ursprüngliche Beweis angepasst werden muss, wird in dieser Arbeit eine ausführliche Herleitung des PORK Algorithmus vorgestellt. Außerdem wird die Integration in adaptive MOR-Methoden wie den stabilitäts-erhaltenden adaptiven rationalen Krylow (SPARK) Algorithmus [30] und das kumulative Reduktionsverfahren (CURE) [42] untersucht.

Als Hauptwerkzeug dieser Arbeit dienen generalisierte Sylvestergleichungen. Es wird gezeigt, dass diese zur Beschreibung von rationalen tangentialen-Eingangs-Krylow Unterräumen genutzt werden können, auch im Falle einer singulären Deskriptormatrix. Außerdem wird eine Formulierung des $\mathcal{H}_2$ inneren Produkts der Übertragungsfunktionen zweier streng properer DAE-Systeme durch projizierte generalisierte Sylvestergleichungen hergleitet. Diese Ergebnisse stellen wichtige Grundlagen für den Beweis von $\mathcal{H}_2$ pseudo-Optimalität durch den PORK Algorithmus dar. Durch ihre Allgemeingültigkeit sind jedoch auch andere Anwendungen denkbar.

Schließlich wird ein effizienter Gesamtalgorithmus zur adaptiven Reduktion von strukturierten linearen DAE-Systemen von beliebigem Index vorgestellt, welcher adaptiv passende Interpolationsdaten und reduzierte Ordnung wählt. Numerische Ergebnisse anhand verschiedener physikalisch motivierter Beispielmodelle zeigen, dass das vorgestellte Verfahren auf typische Problemstellungen in der Technik (unabhängig von Index und Properheit) anwendbar ist. Außerdem wird die Funktionalität auch bei sehr hohem Index anhand eines künstlich erstellten Modells gezeigt.

Acknowledgments

At this point I want to thank all people who supported me during the work on this thesis. First of all I want to thank my advisor Alessandro Castagnotto. Without his extraordinary commitment, be it during the derivation of mathematical relations or the implementation with the sssMOR toolbox, I would definitely not have achieved the presented results. Beside that I want to thank him for the at all times pleasant atmosphere during our meetings.

Moreover I want to thank Prof. Tatjana Stykel, who significantly contributed to this work with her outstanding expertise concerning Gramian based methods and model order reduction in general. I also would like to thank Prof. Boris Lohmann, whose excellent lectures on automatic control and system theory aroused my interest on this topic. Furthermore I thank all contributors (research associates as well as students) to the sssMOR toolbox which made a rapid implementation and test of algorithms possible.

Last but not least I want to thank my family and friends who supported me during this time (and hopefully beyond).

Task Description

The computerized modeling of dynamical systems often results in a system of differential algebraic equations (DAEs), where the state variables are not independent from one another but coupled by the algebraic equations, which implicitly describe a manifold on which the dynamic of the system is constrained. Depending on the dynamics, the constraints and the modeling procedure, DAEs of different indices arise, whereby index 1 DAEs often represent standard electrical circuits, index 2 discretized stokes equations and mechanical systems with non-holonomic constraints and index 3 DAEs mechanical systems with holonomic constraints. The general procedure for the reduction of DAEs requires the computation of spectral projectors onto deflating subspaces, in order to separate the dynamic part from the algebraic, reduce the former while preserving the latter. This procedure is numerically ill-conditioned and not feasible in general. However, if the DAE has a certain structure, then it is possible to identify the dynamic and algebraic part a-priori and adapt the reduction procedures themselves accordingly. Based on the preprint by Castagnotto et al. from 2015, the student shall further investigate Krylov-based reduction of structured DAEs of index 1 to 3. The tasks include a) equivalence of Krylov-Sylvester and extension of the pseudo-optimal rational Krylov (PORK) algorithm b) extension of the cumulative reduction (CURE) procedure and lastly c) extension of the stability-preserving, adaptive rational Krylov (SPARK) algorithm.

Work Program

- Study Krylov based MOR methods in general, with a strong focus on stability-preserving reduction using the CUREd SPARK algorithms.
- Study the general theory on DAEs, their properties and existing reduction procedures. In this process, existing MOR algorithms for DAEs should be implemented in the sssMOR toolbox and used for later benchmarking to the algorithms developed in the thesis.
- Show the equivalence between Krylov and Sylvester for DAEs. Based on this result, extend the PORK algorithm.
- Extend the CURE procedure and the SPARK algorithm to higher index DAEs.

Garching, 30 September 2016

Supervisor

Student

Contents

Glossary

Abbreviations and Acronyms

BT	balanced truncation
CURE	cumulative reduction
DAE	differential algebraic equation
FOM	full order model
HSV	Hankel singular value
IRKA	iterative rational Krylov algorithm
LSE	linear system of equations
LTI	linear time-invariant
MIMO	multiple-input, multiple-output
MOR	model order reduction
ODE	ordinary differential equation
PORK	pseudo-optimal rational Krylov
ROM	reduced order model
SISO	single-input, single-output
SPARK	stability-preserving, adaptive rational Krylov
SVD	singular value decomposition

Notation

Pattern	Meaning	Example
lower case	scalar	λ
bold, lower case	(row or column) vector	$\mathbf{x}$
bold, upper case	matrix	$\mathbf{A}$
bar	complex conjugate	$\overline{\lambda}$
tilde	in Weierstraß canonical form	$\tilde{\mathbf{A}}$, $\tilde{\mathbf{E}}$

Operator	Meaning
$\mathrm{Re}\{\,\dots\}/\mathrm{Im}\{\,\dots\}$	real/imaginary part of a complex variable
$\mathcal{R}(\,\dots)$	image (range) of a matrix
$\dim(\,\dots)$	dimension of a quadratic matrix
$\mathrm{rank}(\,\dots)$	rank of a matrix
$\mathrm{span}\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$	space spanned by the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ...
$\mathrm{colspan}(\,\dots)$	space spanned by the columns of a matrix
$\mathrm{tr}(\,\dots)$	trace of a quadratic matrix
$\det(\,\dots)$	determinant of a quadratic matrix
$\lambda(\,\dots)$	set of (generalized) eigenvalues of a matrix (pair)
$\mathrm{diag}(\mathbf{X}_1, \mathbf{X}_2, \dots)$	block-diagonal matrix consisting of $\mathbf{X}_1$, $\mathbf{X}_2$, ...
$\mathrm{ind}(\,\dots)$	index of a matrix or matrix pencil
$\min/\max\{\,\dots\}$	minimum/maximum of a set or function
$\arg(\,\dots)$	argument, i. e. parameter, of a function
$\mathcal{O}(\,\dots)$	order (polynomial degree)
$\langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{H}_2}$	$\mathcal{H}_2$ inner-product
$\lVert\cdot\rVert_{\mathcal{H}_2}$	$\mathcal{H}_2$ norm

Symbol	Meaning	Example
$\mathbb{N}/\mathbb{R}/\mathbb{C}$	field of natural/real/complex numbers (including 0)	
$\mathbb{F}$	field $\mathbb{R}$ or $\mathbb{C}$	
$\mathcal{G}$	subspace of $\mathcal{H}_2$ according to Definition 4.26	
$\mathcal{H}_2$	Hilbert subspace containing causal, asymptotically stable transfer functions	
$\mathcal{J}$	cost function of the SPARK algorithm	
$\mathcal{K}$	(rational) Krylov subspace	$\mathcal{K}_{\mathrm{ti}}$
$\mathbf{A}$, $\mathbf{E}$	main system matrices of a realization	
$\mathbf{B}$	input matrix of a realization	
$\mathbf{C}$	output matrix of a realization	
$\mathbf{D}$	feedthrough matrix of a realization	
$\mathbf{e}_x$	x-th unit vector	
$f(\,\dots)$	function of ...	
$\mathbf{F}$	parameter matrix of family of transfer functions	
$g(t)/\mathbf{G}(t)$	scalar/matrix-valued impulse response (time domain)	
$\mathbf{G}(s)$	transfer function (frequency domain)	
$\mathbf{H}$	$\mathcal{H}_2$-function	$\mathbf{H}_{\mathrm{opt}}(s)$
$\mathbf{I}_x$	identity matrix of dimension x	$\mathbf{I}_n$, $\mathbf{I}_q$
$\mathbf{J}$	matrix in Jordan canonical form	
$\mathbf{l}/\mathbf{r}$	left/right tangential direction	$\mathbf{l}_{ij}$, $\mathbf{r}_{ij}$
$\mathbf{L}/\mathbf{R}$	interpolation matrix (encoding left/right tangential directions)	

Symbol	Meaning	Example
m	count of system inputs	$\mathbf{u} \in \mathbb{R}^{m\times 1}$
$\mathbf{M}$	moment of a transfer function	$\mathbf{M}_{\mathrm{r}}^{(\mu)}(s_i)$
n	dimension of the full order model	$\mathbf{A} \in \mathbb{R}^{n\times n}$
$\mathbf{N}$	nilpotent matrix	
p	count of system outputs	$\mathbf{y} \in \mathbb{R}^{p\times 1}$
q	dimension of the reduced order model or dimension of the corresponding Krylov subspace	$\mathbf{A}_{\mathrm{r}} \in \mathbb{R}^{q\times q}$ q_i, q_{ij}
$\mathbf{P}/\mathbf{Q}$	left/right transformation matrix for transition into Weierstraß canonical form or improper (polynomial) part of a transfer function	$\tilde{\mathbf{E}} = \mathbf{P}\,\mathbf{E}\,\mathbf{Q}$ $\mathbf{P}(s)$
r	count of tangential directions	r_i
s	expansion point or count of expansion points or frequency (frequency domain)	s_i s $\mathbf{G}(s)$
$\mathbf{S}$	interpolation matrix (encoding expansion points)	$\mathbf{S}_V$, $\mathbf{S}_W$
t	time	$\mathbf{x}(t)$
$\mathbf{T}$	transformation matrix	
$\mathbf{u}$	system input	$\mathbf{u}(t)$
$\mathbf{V}$	basis of an (input) rational Krylov subspace	$\mathbf{V}_{ij}$
$\mathbf{W}$	"left" projection matrix in projective MOR	$\mathbf{W}^{\mathrm{T}}$
$\mathbf{x}$	system state	$\mathbf{x}(t)$
$\mathbf{y}$	system output	$\mathbf{y}(t)$
(α, β)	generalized eigenvalue	
$\delta(t)$	Dirac delta function	
η/ν	index of a matrix/matrix pencil	
θ	proper/improper HSV	θ_i^{p}, θ_i^{im}
$\mathbf{\Gamma}$	controllability/observability Gramian	$\mathbf{\Gamma}_{\mathrm{r}}^{\mathrm{c}}$, $\mathbf{\Gamma}_{\mathrm{r}}^{\mathrm{o}}$
$\mathbf{\Pi}$	spectral projector	$\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_l^\infty$
$\mathbf{\Sigma}$	linear time-invariant (ODE or DAE) system ($\triangleq$ realization + initial value)	
ω	frequency (time domain)	
$\mathbf{\Omega}$	Cholesky factor of a Gramian	$\mathbf{\Omega}^{\mathrm{pc}}$, $\mathbf{\Omega}^{\mathrm{imo}}$
$\imath$	unit imaginary number $\imath = \sqrt{-1}$	

Index	Meaning	Example
H/F/M	related to the H/F/M-system	$\mathbf{A}_{\mathrm{H}}$, $\mathbf{A}_{\mathrm{F}}$, $\mathbf{A}_{\mathrm{M}}$
r	related to the ROM ("reduced" or "reformulated")	$\mathbf{A}_{\mathrm{r}}$, $\mathbf{E}_{\mathrm{r}}$
e	related to the error system	$\mathbf{G}_{\mathrm{e}}$
f	related to the set of finite eigenvalues	$\mathbf{B}_f$, $\mathbf{C}_f$
∞	related to the set of infinite eigenvalues	$\mathbf{B}_\infty$, $\mathbf{C}_\infty$
dyn	related to the dynamical subsystem	n_{dyn}, $\mathbf{I}_{n_{\mathrm{dyn}}}$

Index	Meaning	Example
i	related to the i-th expansion point	$\mathbf{V}_i$, $\mathbf{S}_i$, $\mathbf{R}_i$
j	related to the j-th tangential direction	$\mathbf{V}_{ij}$, $\mathbf{S}_{ij}$ $\mathbf{R}_{ij}$
k	related to the k-th row/column of the corresponding Jordan block	$\mathbf{v}_{ijk}$
l/r	left/right	$\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_r^f$
V/W	related to the basis $\mathbf{V}/\mathbf{W}$	$\mathbf{S}_V$, $\mathbf{S}_W$
bi/bo	block-input/block-output	$\mathcal{K}_{\mathrm{bi}}$, $\mathcal{K}_{\mathrm{bo}}$
ti/to	tangential-input/tangential-output	$\mathcal{K}_{\mathrm{ti}}$, $\mathcal{K}_{\mathrm{to}}$
opt	optimum	$\mathbf{H}_{\mathrm{opt}}$
0	related to the initial state at $t = 0$	$\mathbf{x}_0$
$\perp$	related to the factor $\mathbf{G}_\perp$ of the error model (CURE)	$\mathbf{B}_\perp$, $\mathbf{G}_\perp$
$\mathcal{F}$	related to the factor $\mathbf{G}_\mathcal{F}$ of the error model (CURE)	$\mathbf{G}_\mathcal{F}$
$\flat$	related to the balanced realization of the FOM	$\mathbf{\Gamma}_\flat^{\mathrm{pc}}$, $\mathbf{\Gamma}_\flat^{\mathrm{imo}}$

Superscript	Meaning	Example
> 0	strictly positive	$\mathbb{N}^{>0}$, $\mathbb{R}^{>0}$
(μ)	μ-th derivative with respect to t or s	$\mathbf{G}^{(\mu)}(s)$
T	transpose	$\mathbf{A}^{\mathrm{T}}$
$*$	complex conjugate transpose	$\mathbf{A}^*$
D	Drazin inverse	$\mathbf{E}^{\mathrm{D}}$
P	(related to the) primitive base	$\mathbf{V}^{\mathrm{P}}$, $\mathbf{S}_V^{\mathrm{P}}$, $\mathbf{R}^{\mathrm{P}}$
$\reflectbox{J}$	in mirrored Jordan canonical form	$\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}}$
$\perp$	orthogonal complement of a subspace	$\mathcal{G}^\perp$
c/o	controllability/observability	$\mathbf{\Gamma}_{\mathrm{r}}^{\mathrm{c}}$, $\mathbf{\Gamma}_{\mathrm{r}}^{\mathrm{o}}$
sp	strictly proper	$\mathbf{G}^{\mathrm{sp}}(s)$
p	proper	$\mathbf{\Gamma}^{\mathrm{pc}}$
im	improper	$\mathbf{\Gamma}^{\mathrm{imc}}$

Compound	Meaning
$(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$	LTI (ODE or DAE) system ($\triangleq$ realization + initial value)
$[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$	realization of a transfer function

Chapter 1

Introduction

1.1 Motivation

Since the invention of the first digital computers in the early 1940s, we observe a rapid increase of available computational power. This progress is used to push the limits in industrial and scientific applications: the ability to create large scale models allows to control complex systems, while the refinement of the temporal resolution enables the simulation of high-frequency dynamical effects.

Despite the increasing capabilities, there often is (and maybe will always be) a gap between the desired and the actual performance especially in high-end applications. In such cases one strives to push the available hardware to its limits. Thereby a main obstacle is the computation time: caused by the introduction of multi-core processors in the last decade, the advance in performance affects mainly parallel operations. Since the simulation of a dynamical system is of sequential nature, an increase of the temporal resolution typically involves greater computation times. Another bottleneck are the memory requirements of large dynamical systems: Consider a *dense* matrix of dimension 10^6. Assuming 8 bytes per matrix entry (double-precision), one needs to store 8 terabytes of data.

Although both aspects, computation time and memory consumption, may be acceptable for few offline-simulations on a server-cluster, they are crucial when it comes to a controlling application on an embedded device with limited resources and real-time requirements. Similarly optimization tasks, which require lots of simulations, may also hit the cost-benefit limits, even on a dedicated workstation. In order to handle large dynamical systems while maintaining low computational efforts, methods for model order reduction (MOR) were developed. Using this techniques, it is possible to simplify the original full order model (FOM) in a preceding offline computation step, such that a reduced order model (ROM) with similar behavior can be used to run fast simulations or controlling tasks on real-time hardware.

The main goal of MOR is to obtain a ROM of small dimension, which approximates the behavior of the FOM well, while preserving properties like stability or passivity [2, p. 7]. In order to evaluate the error caused by the reduction, typically the input-output behavior of the FOM and the ROM is compared in some chosen metric (e. g. the $\mathcal{H}_\infty$ or $\mathcal{H}_2$ norm). Numerical efficiency of the reduction process is another objective, otherwise the benefits of MOR compared to simulating the FOM may be lost.

Depending on the type of the dynamical system, different reduction techniques exist. Therefore a classification of the problem has to be made in advance. For this purpose we consider *linear partial differential algebraic equations* which typically connect spacial with temporal state-derivatives of a constrained technical system. In order to solve such systems, a common approach (e. g. during a finite element analysis) is to discretize the equations in space, such that only temporal derivatives remain. After this, one can distinguish between two types of systems: those involving only differential equations, called (linear) ordinary differential equations (ODEs), and those including additional algebraic equations (constraints), called (linear) differential algebraic equations (DAEs).

DAEs occur in a big variety of technical applications like structural and multi-body dynamics, computational electro-magnetics or fluid mechanics [9, p. 2]. Because this work discusses MOR from a mathematical point of view, it does not matter in which context the model was generated. The results in the following chapters concern both, DAEs and ODEs, therefore a short demonstrative example is used to show up the differences: consider the circuit depicted in Figure 1.1, wherein r and c denote (linear) resistors and capacitors respectively. Let the supply voltage u denote the system input, while the voltages x_1 and x_2 compose the system state $\mathbf{x} = [x_1, x_2]^{\mathrm{T}}$ and add up to the output $y = x_1 + x_2$.

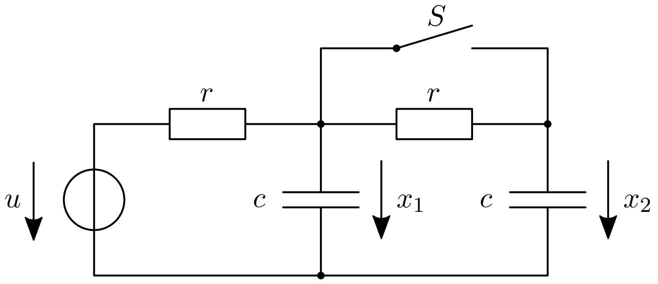


Figure 1.1: Circuit example illustrating the difference between ODEs and DAEs: an open switch S corresponds to an ODE-system, while the closed case can be expressed by a DAE-system.

First assume the switch S to be open. The simple choice $r = c = 1$ allows to model the dynamical system as

$$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}\dot{x}_1\\\dot{x}_2\end{bmatrix} = \begin{bmatrix}-2 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} u\,, \qquad y = \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\,. \tag{1.1}$$

Note that the leading matrix on the left hand side is regular, i. e. $\det(\dots) \neq 0$. Although the system states are coupled, they are not constrained, i. e. they are allowed to have arbitrary values, which is why (1.1) is called an ODE-system.

Now consider the case of a closed switch S, bypassing the right resistor, which is equivalent to adding the constraint $x_1 \overset{!}{=} x_2$. Using again $r = c = 1$ allows to describe the system through

$$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}\dot{x}_1\\\dot{x}_2\end{bmatrix} = \begin{bmatrix}-\frac{1}{2} & 0\\1 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + \begin{bmatrix}\frac{1}{2}\\0\end{bmatrix} u\,, \qquad y = \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\,, \tag{1.2}$$

where the lower part of the left equation is solely algebraic and guarantees $x_1 = x_2$. This time the leading matrix on the left hand side is singular, i. e. $\det(\dots) = 0$. Since we force

$x_1 \overset{!}{=} x_2$, the system states have to be equal *at any time*. This especially holds for the initial state at $t = 0$. Because (1.2) includes both, differential and algebraic equations, it is called a DAE-system.

Note that within this demonstrative example it is possible to find an ODE-formulation even in the case of a closed switch by reformulation of the state space model (known as index reduction). This is because the system has a special property[1], which will be discussed in the following chapter and may not apply to the actual problem.

It is obvious, that the constraint $x_1 = x_2$ has great influence on the behavior of the system since it describes the structure of the circuit. Therefore one should only reduce the complexity related to the differential equations, while keeping the contribution of the algebraic equations unchanged. Thus, the reduction of DAEs introduces an additional goal: reduce the order of the given system, while transferring all constraints to the ROM.

1.2 State of the Art

As mentioned in the previous section, there exist different techniques for MOR. Since this work is related to *linear* dynamical systems, the most popular strategies are singular value decomposition (SVD)-based and Krylov-based methods [4, p. 1096ff.]. One representative of SVD-based MOR is balanced truncation (BT). This method uses a SVD to obtain a *balanced* realization of the FOM, whereby the transformed system states are as “good” controllable as observable. Because states which are “poorly” controllable *and* observable only have weak influence on the input-output behavior of the system, they can be truncated without introducing great approximation error.

The main advantage of BT is, that there exists a global upper error bound, which can be computed a priori (i. e. after the SVD, but before the actual reduction) [2, p. 212]. This allows the user to specify a tolerated approximation error, while the rest of the reduction can be run automatically (i. e. without additional user-interaction). Furthermore stability preservation is guaranteed [4, p. 1114]. A big disadvantage of BT is its computational expense, since two Lyapunov equations of full-order dimension have to be solved, which turns out to be numerically ill-conditioned [2, p. 220]. This makes BT inappropriate for large-scale systems.

Another class of MOR-techniques, rational Krylov subspace methods, which belong to the Krylov-based methods, use a different approach as they try to approximate the transfer function in the frequency domain. Since this is realized by an interpolation through a Laurent series, one speaks of *moment matching* or *rational interpolation* [2, p. 343,346]. Obviously the resulting approximation error depends heavily on the positioning and count of interpolation points. It turns out that the proper choice of expansion points (together with other interpolation data) is a major difficulty.

In comparison to BT, rational Krylov subspace methods are better suited for large-scale systems due to their numerical efficiency [4, p. 1114]. Especially in the case of weakly coupled dynamical systems (which are typical results of a finite element analysis), the sparsity of the system matrices can be exploited. This enables the handling of huge systems (as in the example given in the previous section with 10^6 degrees of freedom),

[1] Since the output $y(t)$ does not depend explicitly on the input $u(t)$, the system described by (1.2) is *strictly proper*, i. e. it can be described by an ODE without feedthrough (see Corollary 2.31).

since sparse matrices occupy much less memory than dense matrices. Another advantage is the freedom in choosing the interpolation parameters: if a frequency domain is of special interest, one can place the interpolation points accordingly while in BT the only user parameter is the desired reduced order or an overall error limit. As already mentioned, this advantage can easily turn into a drawback, since the automatic choice of appropriate interpolation data is difficult. Unfortunately there don't exist universal valid error estimators for rational Krylov subspace methods [4, p. 1098], such that the resulting error remains unknown.[2]

In MOR, one strives to minimize the $\mathcal{H}_\infty$ norm of the error system, since it describes some kind of maximum deviation of the ROM from the FOM. Because it is difficult to deal with the $\mathcal{H}_\infty$ norm especially in the large-scale setting [42, p. 63], the $\mathcal{H}_2$ norm is often used for evaluation instead. As there is currently (up to the author's knowledge) no analytic way to compute a $\mathcal{H}_2$ optimal ROM directly, iterative methods have been investigated which make use of first-order optimality conditions (e. g. in [41]). One of these methods is the well-known iterative rational Krylov algorithm (IRKA) [17, p. 627], where the eigenvalues of the ROM are used to choose the interpolation points of the next iteration step. Although IRKA delivers good results in many cases, convergence and stability-preservation are not guaranteed in the general case. Another drawback is, that the initialization has great influence on the performance [30, p. 49].

A different approach was followed by Thomas Wolf, who discussed the concept of $\mathcal{H}_2$ pseudo-optimality in his dissertation, [42] (earlier published in [43]). On the one hand $\mathcal{H}_2$ pseudo-optimality is a weaker property than $\mathcal{H}_2$ optimality. On the other hand there exists an analytic way to compute the $\mathcal{H}_2$ pseudo-optimal ROM directly (i. e. without iteration) which can be efficiently implemented by the pseudo-optimal rational Krylov (PORK) algorithm [42, p. 91]. Furthermore several *sufficient* conditions for $\mathcal{H}_2$ pseudo-optimality are stated in [42, p. 87f.], which can be used to design similar algorithms. Together with the stability-preserving, adaptive rational Krylov (SPARK) algorithm for single-input, single-output (SISO)-systems introduced by Heiko Panzer in his dissertation, [30] (earlier published in [31]), an iterative scheme, which automatically chooses optimal interpolation points, can be implemented. This way, an (at least local) $\mathcal{H}_2$ optimum can be found.

One advantage of SPARK over IRKA is, that it guarantees an asymptotically stable ROM, since the reduced eigenvalues coincide with the mirrored images of the expansion points (which are chosen in the open right half of the complex plane). Another one is, that it perfectly fits into the cumulative reduction (CURE) framework, presented in [42] (based on [44]), which allows a stepwise assembly of the ROM until the desired reduced order is reached. Note that also IRKA can be combined with CURE as demonstrated in [30, p. 73ff.], but the convergence and stability issues remain.

All mentioned methods are intended to work with ODE-systems. As one has to take special care of the algebraic equations during the reduction of a DAE-system, different variations and modifications of the currently available toolbox arose. The survey carried out in [9] gives a good overview of currently available MOR-techniques for DAE-systems.

First of all SVD-based BT can be used for the reduction of DAEs. As in the ODE-case various types and modifications exist, where *Lyapunov balanced truncation* is the most common variation. The basic idea is to separately reduce the parts of the system

[2] Since the approximation error measured in the $\mathcal{H}_2$ norm includes the transfer function of the original system (full order), an explicit computation may not be feasible.

related to the differential and algebraic equations. For this purpose the so called *spectral projectors* of the FOM have to be known. Note that there are also BT-methods, which do not involve spectral projectors at all, but instead are again limited to small and medium sized problems [9, p. 19].

Concerning rational Krylov subspace methods, IRKA has been adapted in [18, p. B1020] for (local) $\mathcal{H}_2$ optimal reduction of DAEs. Again the FOM is separated into the dynamic and algebraic part, thus the explicit knowledge of the spectral projectors is needed. Other techniques, which do not depend on the explicit computation of spectral projectors, for example those presented in [1] and [18, p. B1020ff.], make use of the special structure of the given problem.

As stated in [9, p. 27], the computation of the spectral projectors is expensive and numerically ill-conditioned especially in the large-scale setting. However, for certain types of DAE-systems, the structure can be exploited to directly obtain analytic expressions. Several examples, including semi-explicit systems of index[3] 1, stokes-like systems of index 2 and mechanical systems of index 1 and 3 are collected in [9, p. 27ff.]. Since the spectral projectors are in general dense matrices, it should be avoided to store them as a whole in the main memory [9, p. 27]. Fortunately they often inherit the block-structure of the system matrices, therefore projector-vector products can be performed block-wise, such that the sparsity of the system matrices is exploited again which leads to an efficient implementation [9, p. 27].

1.3 Tasks, Goals and Assumptions

Since well-established MOR-techniques like BT or IRKA have already been ported to the case of DAE-systems, the focus of this work lies on recently developed methods. For this purpose the already mentioned Krylov-based PORK and SPARK algorithms together with the CURE-framework, introduced by Thomas Wolf, Heiko Panzer and Boris Lohmann few years ago, are analyzed. The main goal of this thesis is to transfer these methods to the DAE-world, thus extending the currently available toolset. The investigations in this contribution are an extension of what can be found in [11] to the general case of (improper) DAEs of arbitrary index and structure.

As the knowledge of the DAE-type significantly reduces the complexity of MOR, only structured problems will be discussed. More precisely, it is assumed, that the spectral projectors are known. Fortunately in technical applications the structure of the FOM is determined by the modeling process (e. g. finite element analysis, modified nodal analysis) and therefore often known in advance. Although the results of this thesis are valid *in theory* for arbitrary linear DAEs, the efficient numerical implementation requires a special structure.

Furthermore only linear time-invariant, first order DAE-systems are considered. Note that every higher order DAE can be reformulated into a system of first order by introducing additional system states, even though this might not be the most efficient way of reduction. Since the structure is considered through spectral projectors, no restrictions regarding the index of the DAE are required. Also the index does not necessarily have to be known in advance (although the modeling process usually admits corresponding

[3] The index of a DAE-system will be introduced in the following chapter. As a short anticipation, a high structural complexity corresponds to a high index.

conclusions). As the focus lies on technical applications, it is assumed that the FOM allows realizations with real-valued system matrices. Moreover only asymptotically stable DAE-systems are considered.

1.4 Outline

In the following chapter necessary fundamentals of DAE-systems are presented. First important concepts like the index, spectral projectors and the Drazin inverse are introduced. Then DAEs are analyzed in the context of control theory, whereby the Weierstraß-canonical form plays an important role. Furthermore the properties *properness*, *controllability* and *observability* are explained.

In Chapter 3 the overall procedure used for reduction is presented. The concept of tangential interpolation is explained, after which rational Krylov subspaces are defined. Then generalized Sylvester equations (especially their solvability and equivalence to rational Krylov subspaces) are discussed, since they play an important role later on. Finally the general framework, which bases on a partitioning into two subsystems, is unveiled.

The main part of this thesis, the reduction of the strictly proper subsystem, is discussed in Chapter 4. The first section deals with the $\mathcal{H}_2$ inner-product of DAEs and presents results, which can be used independently of the rest of this thesis (and even MOR). Then the validity of the PORK and SPARK algorithms (and their integration into the CURE-framework) in the DAE-case is analyzed and proved.

Chapter 5 shows how the improper subsystem can be reformulated to find a minimal realization. For this purpose a short introduction into Lyapunov BT for DAEs according to [9, p. 10] is given. After that the work of Tatjana Stykel in [9] and [39] is used to find a minimal realization of the improper subsystem in an efficient way, i. e. without solving large-scale Lyapunov equations.

In Chapter 6 the complete algorithm proposed in this work is presented, together with all underlying assumptions, such that this chapter can be used as quick reference for the impatient reader. Chapter 7 shows several numerical results, after what final remarks and ideas for future investigations are collected in Chapter 8.

Important note: As this thesis handles proofs in a very general way, an extensive use of indices is inevitable. In order to maintain a precise formulation, special effort has been made to define a proper notation. A comprehensive list can be found right at the beginning of this document.

Due to the subject of this thesis, most results directly apply to DAE-systems. Anyway there are passages, which concern only the special case of ODEs. Note that those parts are used either for clarifications or proofs and do not restrain the results of this work. For better visibility, all theorems and definitions belonging to the ODE-only case are tagged with $\boxed{\textbf{ODE}}$.

Chapter 2

Linear Differential Algebraic Equations

In order to properly reduce DAE-systems, one has to understand their characteristics first. For this purpose, the following sections give a brief introduction into linear DAE-theory. Since the field of DAE-related research is broad, the main focus lies on results used in this thesis. This chapter represents a summary of the theory given in [23] (especially chapter 1 and 2), [26] and [29]. For a more general view on DAE-systems in technical applications, the collections [19], [20] and [21] are recommended.

2.1 Fundamentals

2.1.1 Matrix Pencils

Matrix pencils and pairs are a convenient way of notation while dealing with generalized eigenvalues, Sylvester equations, and DAEs in common:

Definition 2.1. Let $\mathbf{X}$, $\mathbf{Y} \in \mathbb{C}^{u \times v}$. The polynomial matrix $\mathbf{P}(\lambda) = \lambda\mathbf{X} - \mathbf{Y}$ with arbitrary $\lambda \in \mathbb{C}$ is called *(linear) matrix pencil*. An alternative notation is $(\mathbf{X}, \mathbf{Y})$ which is called *matrix pair*.

Within the scope of this thesis only quadratic matrix pencils, i. e. $u = v$, are considered. Note that in the literature different definitions concerning the sign of $\mathbf{Y}$ (e. g. in [26]) or the order of $\mathbf{X}$ and $\mathbf{Y}$ (e. g. in [22]) exist.

The probably most important property of a matrix pencil is *regularity*:

Definition 2.2 (adapted from [23, p. 16])**.** Let $\mathbf{X}$, $\mathbf{Y} \in \mathbb{C}^{u \times v}$. The matrix pencil $\lambda\mathbf{X} - \mathbf{Y}$ is called *regular*, if $u = v$ and $\exists\, \lambda \in \mathbb{C}$ such that $\det(\lambda\mathbf{X} - \mathbf{Y}) \neq 0$. Otherwise it is called *singular*.

If the identity matrix $\mathbf{I}$ is contained in the pair $(\mathbf{X}, \mathbf{Y})$, then an instant classification is possible:

Lemma 2.3. *Let* $\mathbf{X}$, $\mathbf{Y} \in \mathbb{C}^{u \times u}$ *compose the matrix pencil* $\lambda\mathbf{X} - \mathbf{Y}$*. If* $\mathbf{X} = \mathbf{I}_u$ *or* $\mathbf{Y} = \mathbf{I}_u$*, then the matrix pencil* $\lambda\mathbf{X} - \mathbf{Y}$ *is regular.*

Proof. Consider the first case $\mathbf{X} = \mathbf{I}_u$: every $\lambda \in \mathbb{C}$ which does not coincide with the eigenvalues of $\mathbf{Y}$ fulfills $\det(\lambda\,\mathbf{I}_u - \mathbf{Y}) \neq 0$. In the second case $\mathbf{Y} = \mathbf{I}_u$ the choice $\lambda = 0$ guarantees, that $\det(\lambda\,\mathbf{X} - \mathbf{I}_u) = \det(-\mathbf{I}_u) = \pm 1 \neq 0$. ■

The eigenvalues and eigenvectors of a matrix represent its main characteristics. This concept can be extended to the case of a matrix pair, which is known as the *generalized eigenvalue problem*. In contrast to the standard case, a generalized eigenvalue consists of two related scalars[1]:

Definition 2.4 (adapted from [22, p. 68])**.** Let $\mathbf{X},\, \mathbf{Y} \in \mathbb{C}^{u \times u}$. The set of *generalized eigenvalues* of the matrix pair $(\mathbf{X},\, \mathbf{Y})$ is composed of the solutions $(\alpha,\, \beta)$ of

$$\det(\alpha\,\mathbf{X} - \beta\,\mathbf{Y}) = 0\ , \qquad \alpha,\, \beta \in \mathbb{C} \tag{2.1}$$

and is denoted by $\lambda(\mathbf{X},\, \mathbf{Y})$. With the equivalence $\lambda \triangleq \frac{\alpha}{\beta}$, where λ is related to the standard eigenvalue problem, the case $\beta \neq 0$ corresponds to the *finite* eigenvalues $\lambda_f(\mathbf{X},\, \mathbf{Y})$ and the case $\beta = 0$ corresponds to the *infinite* eigenvalues $\lambda_\infty(\mathbf{X},\, \mathbf{Y})$.

Furthermore every non-trivial vector $\mathbf{z}_r \in \mathbb{C}^u$ is called *right generalized eigenvector* of the pair $(\mathbf{X},\, \mathbf{Y})$, if it satisfies $\alpha\,\mathbf{X}\,\mathbf{z}_r = \beta\,\mathbf{Y}\,\mathbf{z}_r$. Accordingly every non-trivial vector $\mathbf{z}_l \in \mathbb{C}^u$ is called *left generalized eigenvector* of the pair $(\mathbf{X},\, \mathbf{Y})$, if it satisfies $\alpha\,\mathbf{z}_l^*\,\mathbf{X} = \beta\,\mathbf{z}_l^*\,\mathbf{Y}$.

In the following the term *generalized eigenvalue* will be used for both, a description by a scalar ($\lambda = \frac{\alpha}{\beta}$) and a pair of scalars $(\alpha,\, \beta)$, since both notations are equivalent.

According to Definition 2.4, there may be generalized eigenvalues at infinity (i. e. for $\beta = 0$). In the context of DAEs, those correspond to the algebraic part, what can be considered as an *infinitely* fast dynamical subsystem. Using the properties of equivalent matrices, it is possible to find a partitioning corresponding to the finite and infinite eigenvalues:[2]

Lemma 2.5 (adapted from [23, pp. 13,16])**.** *Let* $\mathbf{E},\, \mathbf{A} \in \mathbb{C}^{n \times n}$*. If the matrix pencil* $\lambda\,\mathbf{E} - \mathbf{A}$ *is regular, then there exist regular transformation matrices* $\mathbf{P},\, \mathbf{Q} \in \mathbb{C}^{n \times n}$ *such that*

$$\tilde{\mathbf{E}} = \mathbf{P}\,\mathbf{E}\,\mathbf{Q} = \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{N} \end{bmatrix}, \quad \tilde{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix} \tag{2.2}$$

where both $\mathbf{J}$ *and* $\mathbf{N}$ *are in Jordan-canonical form. Furthermore* $\mathbf{N}$ *is nilpotent of index* ν*, i. e.* $\mathbf{N}^\nu = \mathbf{0}$ *and* $\mathbf{N}^{\nu-1} \neq \mathbf{0}$*, and the diagonal elements of* $\mathbf{J}$ *coincide with* $\lambda_f(\mathbf{E},\, \mathbf{A})$ *[29, p. 409]. The identity matrices* $\mathbf{I}_{n_f}$ *and* $\mathbf{I}_{n_\infty}$ *are of dimension* n_f *and* n_∞ *which denote the count of finite and infinite generalized eigenvalues of the matrix pair* $(\mathbf{E},\, \mathbf{A})$ *respectively.*

Note that Lemma 2.5 does not provide any information about the calculation of $\mathbf{P}$ and $\mathbf{Q}$. Within the scope of this work, both transformation matrices are only of theoretical interest because they are not needed in the implementation afterwards. Thus their existence is sufficient.

Beside regularity the *index* of a matrix pencil is another important property, since it is a measure for the structural complexity of the related DAE-system:

[1] Note that a slightly different notation is used in comparison with [22] as the scalars α and β are interchanged.

[2] In the literature (e. g. [23]), the special form of $(\mathbf{E},\, \mathbf{A})$ given in (2.2) is called *Weierstraß canonical form* of the matrix pencil. Within the scope of this thesis, this term will instead be associated with the corresponding LTI DAE-system introduced in Section 2.2.

Definition 2.6 (adapted from [23, p. 18])**.** Let $\lambda\mathbf{E} - \mathbf{A}$ be a regular matrix pencil and $\lambda\tilde{\mathbf{E}} - \tilde{\mathbf{A}}$ denote its transformation according to Lemma 2.5. Then the index of nilpotency ν of $\mathbf{N}$ in (2.2) is called the *index of the matrix pencil* $\lambda\mathbf{E} - \mathbf{A}$ and is denoted by $\nu = \operatorname{ind}(\lambda\mathbf{E} - \mathbf{A})$.

The index of a (single) matrix is defined as a special case of Definition 2.6:

Definition 2.7 (adapted from [23, p. 24])**.** Let $\mathbf{X} \in \mathbb{C}^{u\times u}$. Then the index of the matrix pencil $\lambda\mathbf{X} - \mathbf{I}_u$ is called *index of* $\mathbf{X}$ and is denoted[3] by $\eta = \operatorname{ind}(\mathbf{X})$.

2.1.2 Spectral Projectors

In order to separate dynamic and algebraic contributions of a DAE-system, *spectral projectors* can be used. They play a key role in this thesis, since structured problems are considered, for which analytic expressions of the spectral projectors can be found in several applications [9, p. 27].

Definition 2.8 ([29, p. 409])**.** Let $\mathbf{P}$ and $\mathbf{Q}$ denote the transformation matrices related to $\lambda\mathbf{E} - \mathbf{A}$ according to Lemma 2.5. The matrices

$$\mathbf{\Pi}_l^f = \mathbf{P}^{-1}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P} \qquad \text{and} \qquad \mathbf{\Pi}_r^f = \mathbf{Q}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1} \tag{2.3}$$

are called *spectral projectors* onto the *left* and *right* deflating subspace of $\lambda\mathbf{E} - \mathbf{A}$ corresponding to the *finite* eigenvalues. The matrices

$$\mathbf{\Pi}_l^\infty = \mathbf{P}^{-1}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_\infty}\end{bmatrix}\mathbf{P} \qquad \text{and} \qquad \mathbf{\Pi}_r^\infty = \mathbf{Q}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_\infty}\end{bmatrix}\mathbf{Q}^{-1} \tag{2.4}$$

are called *spectral projectors* onto the *left* and *right* deflating subspace of $\lambda\mathbf{E} - \mathbf{A}$ corresponding to the *infinite* eigenvalues.

Again, the relations in (2.3) and (2.4) are only of theoretical interest, since $\mathbf{P}$ and $\mathbf{Q}$ are unknown. Instead it is assumed, that explicit formulas for the computation of the spectral projectors are available. Considering that, expressions for $\mathbf{\Pi}_l^f$ and $\mathbf{\Pi}_r^f$ are sufficient, because the respective counterparts $\mathbf{\Pi}_l^\infty$ and $\mathbf{\Pi}_r^\infty$ are determined by a simple relationship:

Lemma 2.9. *Let* $\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_r^f$, $\mathbf{\Pi}_l^\infty$ *and* $\mathbf{\Pi}_r^\infty$ *be spectral projectors according to Definition 2.8. Then the following relations hold:*

$$\mathbf{\Pi}_l^f + \mathbf{\Pi}_l^\infty = \mathbf{I}_n\,, \qquad \mathbf{\Pi}_r^f + \mathbf{\Pi}_r^\infty = \mathbf{I}_n\,. \tag{2.5}$$

Proof. The proof directly follows from inserting (2.3) and (2.4) into (2.5). ■

[3] Note that a different symbol (η instead of ν) is used, in order to avoid ambivalence.

2.1.3 The Drazin Inverse

The introductory example in Chapter 1 shows, that the algebraic equations of a DAE-system cause singularity of a specific matrix (see (1.2)). Unfortunately several results from basic linear system theory (e. g. explicit solution formulas) require regularity of this matrix. In order to find similar relations for the DAE-case, the *Drazin inverse*, which is a generalization of the matrix inverse, can be used:

Definition 2.10 ([23, p. 24])**.** Every $\mathbf{X} \in \mathbb{C}^{u\times u}$ has one and only one Drazin inverse $\mathbf{X}^{\mathrm{D}}$ which is defined through

$$\begin{aligned} \mathbf{X}\,\mathbf{X}^{\mathrm{D}} &= \mathbf{X}^{\mathrm{D}}\,\mathbf{X}\,, \\ \mathbf{X}^{\mathrm{D}}\,\mathbf{X}\,\mathbf{X}^{\mathrm{D}} &= \mathbf{X}^{\mathrm{D}}\,, \\ \mathbf{X}^{\mathrm{D}}\,\mathbf{X}^{\eta+1} &= \mathbf{X}^{\eta}\,, \qquad \eta = \mathrm{ind}(\mathbf{X})\,. \end{aligned} \tag{2.6}$$

Note that the index of the matrix $\mathbf{X}$, denoted by η, is essential for the definition of $\mathbf{X}^{\mathrm{D}}$. As a true generalization, the Drazin inverse complies with the case of a regular matrix:

Lemma 2.11 ([23, p. 25])**.** *The Drazin inverse of a regular matrix* $\mathbf{X} \in \mathbb{C}^{u\times u}$ *is equal to its inverse* $\mathbf{X}^{-1}$*, i. e.* $\mathbf{X}^{\mathrm{D}} = \mathbf{X}^{-1}$ *if* $\det(\mathbf{X}) \neq 0$.

In the following, Lemma 2.12 and Lemma 2.13 present important properties concerning nilpotent and blockdiagonal matrices, which will be exploited in the following section.

Lemma 2.12. *The Drazin inverse of a nilpotent matrix* $\mathbf{X} \in \mathbb{C}^{u\times u}$ *with index of nilpotency* η*, i. e.* $\mathbf{X}^{\eta} = \mathbf{0}$ *and* $\mathbf{X}^{\eta-1} \neq \mathbf{0}$*, is the zero-matrix.*

Proof. To prove the statement, assume that $\mathbf{X}^{\mathrm{D}} = \mathbf{0}$ holds. Inserting $\mathbf{X}^{\mathrm{D}}$ into (2.6) shows, that all three conditions

$$\begin{aligned} \mathbf{X}\,\mathbf{0} &= \mathbf{0}\,\mathbf{X}\,, \\ \mathbf{0}\,\mathbf{X}\,\mathbf{0} &= \mathbf{0}\,, \\ \mathbf{0}\,\mathbf{X}^{\eta+1} &= \mathbf{X}^{\eta} = \mathbf{0} \end{aligned} \tag{2.7}$$

are satisfied, which leads to the conclusion, that $\mathbf{0}$ is indeed the Drazin inverse of $\mathbf{X}$. Note that the Drazin inverse is *unique* according to Definition 2.10. ■

Lemma 2.13. *Let* $\mathbf{Z} \in \mathbb{C}^{w\times w}$ *be blockdiagonal consisting of* $\mathbf{X} \in \mathbb{C}^{u\times u}$ *with* $\eta_x = \mathrm{ind}(\mathbf{X})$ *and* $\mathbf{Y} \in \mathbb{C}^{v\times v}$ *with* $\eta_y = \mathrm{ind}(\mathbf{Y})$*, i. e.* $\mathbf{Z} = \mathrm{diag}(\mathbf{X},\, \mathbf{Y})$*. Then*

(i) the index of $\mathbf{Z}$ *is given by* $\eta_z = \mathrm{ind}(\mathbf{Z}) = \max(\eta_x,\, \eta_y)$ *and*

(ii) the Drazin inverse of $\mathbf{Z}$ *is* $\mathbf{Z}^{\mathrm{D}} = \mathrm{diag}(\mathbf{X}^{\mathrm{D}},\, \mathbf{Y}^{\mathrm{D}})$*, where* $\mathbf{X}^{\mathrm{D}}$ *and* $\mathbf{Y}^{\mathrm{D}}$ *denote the Drazin inverses of* $\mathbf{X}$ *and* $\mathbf{Y}$ *respectively.*

Proof. The proof is contained in Appendix A. ■

2.2 Descriptor Systems in Control Theory

Because linear control theory is well understood and (comparatively) pleasant to deal with, most technical applications use linear models of the real system. Even if the underlying physical relationships are nonlinear, it is often possible to approximate the dynamical behavior by a linear model as long as the system state remains in a specified operating range. For this reason, all following investigations are restricted to linear time-invariant (LTI) DAE-systems alias *descriptor systems*:

Definition 2.14. The system

$$\mathbf{E}\,\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{B}\,\mathbf{u}(t)\ , \qquad \mathbf{y}(t) = \mathbf{C}\,\mathbf{x}(t)\ , \qquad \mathbf{x}(t=0) = \mathbf{x}_0 \tag{2.8}$$

with state $\mathbf{x}(t) \in \mathbb{R}^n$, input $\mathbf{u}(t) \in \mathbb{R}^m$, output $\mathbf{y}(t) \in \mathbb{R}^p$ and constant system matrices $\mathbf{E} \in \mathbb{R}^{n\times n}$, $\det(\mathbf{E}) = 0$, $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$, $\mathbf{C} \in \mathbb{R}^{p\times n}$, is called *LTI DAE-system* or *descriptor system* and is abbreviated by $\mathbf{\Sigma} = (\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C},\,\mathbf{x}_0)$.

Remember the introductory example given in Chapter 1, which has the same layout as (2.8): because DAE-systems are considered, algebraic equations are involved, which result in a singular matrix $\mathbf{E}$ (thus $\det(\mathbf{E}) = 0$ in Definition 2.14). Therefore the singularity of $\mathbf{E}$ is the main criterion for distinguishing DAEs from ODEs.

Note that adding a feedthrough by $\mathbf{y}(t) = \mathbf{C}\,\mathbf{x}(t) + \mathbf{D}\,\mathbf{u}(t)$ would result in a more general form of (2.8). Since $\mathbf{D}$ affects the solution $\mathbf{y}(t)$ via a summation, it is always possible to neglect the term $\mathbf{D}\,\mathbf{u}(t)$ during MOR and append it to the ROM afterwards. In order to keep things simple, it is supposed (without loss of generality) that $\mathbf{D} = \mathbf{0}$ holds. Moreover the system states and matrices are assumed to be real-valued, since most technical systems relate real-valued input- and output-variables.

As stated in the previous section, the most important property of the matrix pencil $\lambda\,\mathbf{E} - \mathbf{A}$ is its regularity[4]. This is because it has direct influence on the solvability of the related DAE-system:

Theorem 2.15 ([23, p. 16])**.** *The DAE-system* $(\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C},\,\mathbf{x}_0)$ *is* solvable *with* unique *solution* $\mathbf{y}(t)$*, if and only if the matrix pencil* $\lambda\,\mathbf{E} - \mathbf{A}$ *is regular and the initial state* $\mathbf{x}_0$ *is consistent, i. e. it lies on the constraint manifold described by the algebraic part of the DAE.*

Since the goal of this thesis is to provide methods for MOR in usual technical applications, it is assumed, that $\lambda\,\mathbf{E} - \mathbf{A}$ is regular and the initial state $\mathbf{x}_0$ is consistent, i. e. the DAE-systems has a *unique* solution.

2.2.1 The Weierstraß Canonical Form

The Weierstraß canonical form of a DAE-system is equivalent to the transformation of an ODE-system into modal coordinates. Because it separates the dynamic and algebraic contributions of the input-output behavior, it is a perfectly suited tool for the development of model reduction theory.

[4] Note that $\det(\mathbf{E}) = 0$ does not hold any information about the regularity of $\lambda\,\mathbf{E} - \mathbf{A}$.

Definition 2.16 ([23, p. 17])**.** Consider the DAE-system $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$ and the transformation matrices $\mathbf{P}$, $\mathbf{Q}$ according to Lemma 2.5. Multiplying (2.8) with $\mathbf{P}$ from the left leads together with $\mathbf{Q}\,\tilde{\mathbf{x}}(t) = \mathbf{x}(t)$ to the notation

$$\begin{aligned}
\underbrace{\begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{N} \end{bmatrix}}_{\tilde{\mathbf{E}}} \underbrace{\begin{bmatrix} \dot{\tilde{\mathbf{x}}}_f(t) \\ \dot{\tilde{\mathbf{x}}}_\infty(t) \end{bmatrix}}_{\dot{\tilde{\mathbf{x}}}(t)} &= \underbrace{\begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}}_{\tilde{\mathbf{A}}} \underbrace{\begin{bmatrix} \tilde{\mathbf{x}}_f(t) \\ \tilde{\mathbf{x}}_\infty(t) \end{bmatrix}}_{\tilde{\mathbf{x}}(t)} + \underbrace{\begin{bmatrix} \tilde{\mathbf{B}}_f \\ \tilde{\mathbf{B}}_\infty \end{bmatrix}}_{\tilde{\mathbf{B}}} \mathbf{u}(t)\ , \\
\mathbf{y}(t) &= \underbrace{\begin{bmatrix} \tilde{\mathbf{C}}_f & \tilde{\mathbf{C}}_\infty \end{bmatrix}}_{\tilde{\mathbf{C}}} \underbrace{\begin{bmatrix} \tilde{\mathbf{x}}_f(t) \\ \tilde{\mathbf{x}}_\infty(t) \end{bmatrix}}_{\tilde{\mathbf{x}}(t)}\ ,
\end{aligned} \tag{2.9}$$

which is called *Weierstraß canonical form* of the DAE-system.

The formulation (2.9) allows to split the DAE into a *slow* and a *fast* subsystem corresponding to the finite and infinite eigenvalues respectively:

$$\begin{aligned}
\dot{\tilde{\mathbf{x}}}_f(t) &= \mathbf{J}\,\tilde{\mathbf{x}}_f(t) + \tilde{\mathbf{B}}_f\,\mathbf{u}(t)\ , && \text{(slow subsystem)} \\
\mathbf{N}\,\dot{\tilde{\mathbf{x}}}_\infty(t) &= \tilde{\mathbf{x}}_\infty(t) + \tilde{\mathbf{B}}_\infty\,\mathbf{u}(t)\ , && \text{(fast subsystem)}
\end{aligned} \tag{2.10}$$

where the output can be computed through superposition of both contributions:

$$\mathbf{y}(t) = \tilde{\mathbf{C}}_f\,\tilde{\mathbf{x}}_f(t) + \tilde{\mathbf{C}}_\infty\,\tilde{\mathbf{x}}_\infty(t)\ . \tag{2.11}$$

Since the slow subsystem is of ODE-type (it contains only finite eigenvalues), one can write its explicit solution [46, p. 468ff.]:

$$\tilde{\mathbf{x}}_f(t) = e^{\mathbf{J}(t-t_0)}\,\tilde{\mathbf{x}}_{f0} + \int_{t_0}^{t} e^{\mathbf{J}(t-\tau)}\,\tilde{\mathbf{B}}_f\,\mathbf{u}(\tau)\,\mathrm{d}\tau\ . \tag{2.12}$$

The explicit solution of the fast subsystem can be obtained through several derivation steps [26, p. 86]:

$$\begin{aligned}
\tilde{\mathbf{x}}_\infty(t) &= \mathbf{N}\,\dot{\tilde{\mathbf{x}}}_\infty(t) - \tilde{\mathbf{B}}_\infty\,\mathbf{u}(t) && \left|\ \frac{\mathrm{d}}{\mathrm{d}t}(\dots)\right. \\
\overbrace{\dot{\tilde{\mathbf{x}}}_\infty(t)} &= \mathbf{N}\,\ddot{\tilde{\mathbf{x}}}_\infty(t) - \tilde{\mathbf{B}}_\infty\,\dot{\mathbf{u}}(t) && \left|\ \frac{\mathrm{d}}{\mathrm{d}t}(\dots)\right. \\
\overbrace{\ddot{\tilde{\mathbf{x}}}_\infty(t)} &= \mathbf{N}\,\tilde{\mathbf{x}}_\infty^{(3)}(t) - \tilde{\mathbf{B}}_\infty\,\ddot{\mathbf{u}}(t) && \left|\ \frac{\mathrm{d}}{\mathrm{d}t}(\dots)\right. \\
&\ddots
\end{aligned} \tag{2.13}$$

which leads after $\nu - 1$ steps to

$$\tilde{\mathbf{x}}_\infty(t) = \mathbf{N}^\nu\,\tilde{\mathbf{x}}_\infty^{(\nu)}(t) - \sum_{w=0}^{\nu-1} \mathbf{N}^w\,\tilde{\mathbf{B}}_\infty\,\mathbf{u}^{(w)}(t)\ . \tag{2.14}$$

Exploiting the nilpotency of $\mathbf{N}$, i. e. $\mathbf{N}^\nu = \mathbf{0}$, results in

$$\tilde{\mathbf{x}}_\infty(t) = -\sum_{w=0}^{\nu-1} \mathbf{N}^w\,\tilde{\mathbf{B}}_\infty\,\mathbf{u}^{(w)}(t)\ . \tag{2.15}$$

The explicit solution of the overall system can therefore be written as

$$\mathbf{y}(t) = \tilde{\mathbf{C}}_f \, e^{\mathbf{J}(t-t_0)} \, \tilde{\mathbf{x}}_{f0} + \tilde{\mathbf{C}}_f \int_{t_0}^{t} e^{\mathbf{J}(t-\tau)} \, \tilde{\mathbf{B}}_f \, \mathbf{u}(\tau) \, \mathrm{d}\tau - \tilde{\mathbf{C}}_\infty \sum_{w=0}^{\nu-1} \mathbf{N}^w \, \tilde{\mathbf{B}}_\infty \, \mathbf{u}^{(w)}(t) \; . \tag{2.16}$$

Through (2.16) it is evident, that the parameter ν has great influence on the characteristics of the solution and is therefore a major criterion for the classification of DAE-systems:

Definition 2.17 ([26, p. 86])**.** The index ν of the matrix pencil $\lambda\,\mathbf{E} - \mathbf{A}$ is called *(differentiation) index*[5] of the DAE-system.

Regarding the modeling process, the differentiation index specifies the minimum count of derivation steps needed, in order to reformulate a DAE- into an ODE-system [23, p. 7]. Put simply, it describes how "far away" a DAE is from an ODE, thus measuring the complexity of the problem.

Because algebraic constraints have to be fulfilled at any time, special care has to be taken while defining the initial state $\mathbf{x}_0$, which decomposes into $\tilde{\mathbf{x}}_{f0}$ related to the slow subsystem and $\tilde{\mathbf{x}}_{\infty 0}$ related to the fast subsystem:

$$\mathbf{x}_0 = \mathbf{Q} \begin{bmatrix} \tilde{\mathbf{x}}_{f0} \\ \tilde{\mathbf{x}}_{\infty 0} \end{bmatrix} \; . \tag{2.17}$$

While $\tilde{\mathbf{x}}_{f0}$ is arbitrary, $\tilde{\mathbf{x}}_{\infty 0}$ has to fulfill

$$\tilde{\mathbf{x}}_{\infty 0} = \tilde{\mathbf{x}}_\infty(t_0 = 0) = -\sum_{w=0}^{\nu-1} \mathbf{N}^w \, \tilde{\mathbf{B}}_\infty \, \mathbf{u}^{(w)}(0) \tag{2.18}$$

in order to be consistent [26, p. 86]. As mentioned above, $\mathbf{x}_0$ and thus $\tilde{\mathbf{x}}_{\infty 0}$ are assumed to be consistent.

Beside the separation of the slow and fast subsystem, the Weierstraß canonical form (or more precisely the distinction between finite and infinite eigenvalues) allows to make a statement about the stability of the system:

Definition 2.18 (adapted from [29, p. 409] and [40, p. 844])**.** The (autonomous) unperturbed[6] DAE-system $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ is called *asymptotically stable*, if and only if the finite generalized eigenvalues of the matrix pair $(\mathbf{E},\, \mathbf{A})$ lie in the open left half of the complex plane, i. e. $\mathrm{Re}\,\{\lambda_f(\mathbf{E},\, \mathbf{A})\} < 0$. Furthermore the matrix pencil $\lambda\,\mathbf{E} - \mathbf{A}$ is called *c-stable*.

Note that although infinite eigenvalues introduce several obstacles into MOR, they at least have no influence on the stability of the system. As the FOM is supposed to be asymptotically stable, it is assumed, that all finite eigenvalues lie in the open left half of the complex plane. This will be of major importance in the following chapters.

[5] Note that there exist several other index concepts like the *perturbation* and *strangeness* index [23, p. 6f.], which will not be discussed.

[6] In the case of a perturbation of the matrices **E** and **A** one has to take special care during stability analysis (see "Robust Stability of Differential-Algebraic Equations" in [19, p. 63ff.]). In the following only the nominal (i. e. unperturbed) system is considered.

2.2.2 Transfer Function and Properness

The transfer function of a DAE-system $(\mathbf{E},\ \mathbf{A},\ \mathbf{B},\ \mathbf{C},\ \mathbf{x}_0)$ reads as

$$\mathbf{G}(s) = \mathbf{C}\,(s\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{B}\,, \tag{2.19}$$

which is exactly the same as in the ODE-case. Since a different combination of system matrices may generate the same transfer function, the term "realization" is introduced:

Definition 2.19. A set of matrices $\mathbf{E}$, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ which lead to the transfer function $\mathbf{G}(s) = \mathbf{C}\,(s\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{B}$ is called *realization of* $\mathbf{G}(s)$ and is denoted by $[\mathbf{E},\ \mathbf{A},\ \mathbf{B},\ \mathbf{C}]$. Every transfer function has infinitely many realizations, which are called *restricted system equivalent* [27, p. 637] (to each other).

Considering ODE-systems, $\mathbf{y}(t)$ does not *explicitly* depend on $\mathbf{u}(t)$ (assumed that $\mathbf{D} = \mathbf{0}$ holds), but rather in an indirect way through integration. This may not be the case for DAEs: as (2.16) shows, $\mathbf{y}(t)$ contains explicit expressions of $\mathbf{u}(t)$ and even its derivatives depending on the interaction of $\tilde{\mathbf{C}}_\infty$, $\mathbf{N}$ and $\tilde{\mathbf{B}}_\infty$. The way how $\mathbf{u}(t)$ influences $\mathbf{y}(t)$ is described by the property of *properness*:

Definition 2.20 ([26, p. 87])**.** A DAE-system $(\mathbf{E},\ \mathbf{A},\ \mathbf{B},\ \mathbf{C},\ \mathbf{x}_0)$ is called

- *proper*, if its output vector $\mathbf{y}(t)$ does not explicitly depend on the derivatives of the input vector $(\dot{\mathbf{u}}(t),\ \ddot{\mathbf{u}}(t),\ \ldots)$, i. e.

$$\mathbf{y}(t) \neq f\left(\mathbf{u}^{(w)}(t)\right)\,, \qquad \forall\, w \in \mathbb{N}^{>0}\,, \tag{2.20}$$

- *strictly proper*, if its output vector $\mathbf{y}(t)$ does not explicitly depend on the input vector $\mathbf{u}(t)$ or one of its derivatives $(\dot{\mathbf{u}}(t),\ \ddot{\mathbf{u}}(t),\ \ldots)$, i. e.

$$\mathbf{y}(t) \neq f\left(\mathbf{u}^{(w)}(t)\right)\,, \qquad \forall\, w \in \mathbb{N}^{\geq 0}\,, \tag{2.21}$$

- *improper*, if it is neither strictly proper, nor proper.

Analyzing the product of $\tilde{\mathbf{C}}_\infty$, $\mathbf{N}$ and $\tilde{\mathbf{B}}_\infty$ one can formulate criteria for properness and strictly properness:

Lemma 2.21. *Let* $[\tilde{\mathbf{E}},\ \tilde{\mathbf{A}},\ \tilde{\mathbf{B}},\ \tilde{\mathbf{C}}]$ *be a realization of* $\mathbf{G}(s)$ *in Weierstraß canonical form. Then the DAE-system is*

- *proper, if and only if*

$$\tilde{\mathbf{C}}_\infty\,\mathbf{N}^w\,\tilde{\mathbf{B}}_\infty = \mathbf{0}\,, \qquad \forall\, w = 1,\, \ldots,\, \nu - 1\,. \tag{2.22}$$

holds and

- *strictly proper, if and only if*

$$\tilde{\mathbf{C}}_\infty\,\mathbf{N}^w\,\tilde{\mathbf{B}}_\infty = \mathbf{0}\,, \qquad \forall\, w = 0,\, \ldots,\, \nu - 1\,. \tag{2.23}$$

holds.

Proof. Since (2.20) has to hold for any t and the derivatives of $\mathbf{u}(t)$ are arbitrary at a specific t, (2.22) is a necessary and sufficient condition in order that all terms containing $\dot{\mathbf{u}}(t)$, $\ddot{\mathbf{u}}(t)$, ..., $\mathbf{u}^{(\nu-1)}(t)$ in (2.16) vanish. The proof regarding the criterion for strictly properness is analogous. ■

Using this classification, the solution of a general DAE-system (2.16) can be split up into a strictly proper part

$$\mathbf{y}^{\mathrm{sp}}(t) = \tilde{\mathbf{C}}_f \, e^{\mathbf{J}(t-t_0)} \, \tilde{\mathbf{x}}_{f0} + \tilde{\mathbf{C}}_f \int_{t_0}^{t} e^{\mathbf{J}(t-\tau)} \, \tilde{\mathbf{B}}_f \, \mathbf{u}(\tau) \, \mathrm{d}\tau \tag{2.24}$$

and an improper part

$$\mathbf{y}^{\mathrm{im}}(t) = -\tilde{\mathbf{C}}_\infty \sum_{w=0}^{\nu-1} \mathbf{N}^w \, \tilde{\mathbf{B}}_\infty \, \mathbf{u}^{(w)}(t) \; , \tag{2.25}$$

which add up to the overall solution $\mathbf{y}(t) = \mathbf{y}^{\mathrm{sp}}(t) + \mathbf{y}^{\mathrm{im}}(t)$.

Since the transfer function is invariant under state space transformations, it can also be written using the Weierstraß canonical form:

$$\begin{aligned}
\mathbf{G}(s) &= \mathbf{C} \, (s \, \mathbf{E} - \mathbf{A})^{-1} \, \mathbf{B} = \tilde{\mathbf{C}} \left(s \, \tilde{\mathbf{E}} - \tilde{\mathbf{A}} \right)^{-1} \tilde{\mathbf{B}} \\
&= \begin{bmatrix} \tilde{\mathbf{C}}_f & \tilde{\mathbf{C}}_\infty \end{bmatrix} \begin{bmatrix} s \, \mathbf{I}_{n_f} - \mathbf{J} & \mathbf{0} \\ \mathbf{0} & s \, \mathbf{N} - \mathbf{I}_{n_\infty} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{B}}_f \\ \tilde{\mathbf{B}}_\infty \end{bmatrix} \\
&= \begin{bmatrix} \tilde{\mathbf{C}}_f & \tilde{\mathbf{C}}_\infty \end{bmatrix} \begin{bmatrix} \left(s \, \mathbf{I}_{n_f} - \mathbf{J} \right)^{-1} & \mathbf{0} \\ \mathbf{0} & (s \, \mathbf{N} - \mathbf{I}_{n_\infty})^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{B}}_f \\ \tilde{\mathbf{B}}_\infty \end{bmatrix} \\
&= \tilde{\mathbf{C}}_f \left(s \, \mathbf{I}_{n_f} - \mathbf{J} \right)^{-1} \tilde{\mathbf{B}}_f + \tilde{\mathbf{C}}_\infty \, (s \, \mathbf{N} - \mathbf{I}_{n_\infty})^{-1} \, \tilde{\mathbf{B}}_\infty \; .
\end{aligned} \tag{2.26}$$

This again allows the separation of the strictly proper part

$$\mathbf{G}^{\mathrm{sp}}(s) = \tilde{\mathbf{C}}_f \left(s \, \mathbf{I}_{n_f} - \mathbf{J} \right)^{-1} \tilde{\mathbf{B}}_f \tag{2.27}$$

corresponding to the strictly proper part of the solution $\mathbf{y}^{\mathrm{sp}}(t)$ and an improper part

$$\mathbf{P}(s) = \tilde{\mathbf{C}}_\infty \, (s \, \mathbf{N} - \mathbf{I}_{n_\infty})^{-1} \, \tilde{\mathbf{B}}_\infty \overset{\text{(Neumann series)}}{=} -\tilde{\mathbf{C}}_\infty \sum_{w=0}^{\nu-1} \mathbf{N}^w \, \tilde{\mathbf{B}}_\infty \, s^w \tag{2.28}$$

corresponding to the improper part of the solution $\mathbf{y}^{\mathrm{im}}(t)$. While $\mathbf{G}^{\mathrm{sp}}(s)$ is a rational function with $\mathcal{O}(\text{numerator}) < \mathcal{O}(\text{denominator})$, $\mathbf{P}(s)$ is a polynomial of order $\mathcal{O}(\mathbf{P}(s)) \leq \nu - 1$.

The different appearance of strictly proper, proper and improper DAE-systems in the frequency response diagram is shown in Figure 2.1. For simplicity the transfer functions

$$\begin{aligned}
\mathbf{G}_1(s) &= \frac{2}{s+1} \; , & &\Rightarrow \quad \text{strictly proper} \\
\mathbf{G}_2(s) &= \frac{1}{s+1} + 1 \; , & &\Rightarrow \quad \text{proper} \\
\mathbf{G}_3(s) &= \frac{1}{s+1} + 1 + s \; , & &\Rightarrow \quad \text{improper}
\end{aligned} \tag{2.29}$$

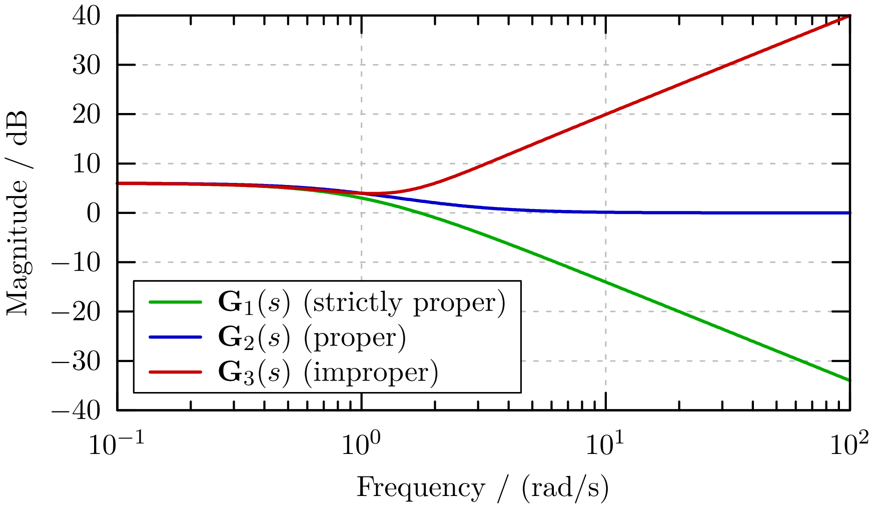


Figure 2.1: Frequency response in dependency of properness: the amplitude of the strictly proper transfer function drops for high frequencies, while it tends to a constant value in the proper case. In contrast, the term s in the improper $\mathbf{G}_3(s)$ leads to unbounded amplification.

are used. As one can see, the graphs diverge with increasing excitation frequency. For $\omega \to \infty$ the improper frequency response tends to infinity, while it stays bounded in the proper and strictly proper case.

Because the transformation matrices $\mathbf{P}$ and $\mathbf{Q}$ and thus the Weierstraß canonical form of the DAE-system are usually unknown, the partitioning of the transfer function according to (2.27) and (2.28) might not be helpful. Instead the spectral projectors can be used to accomplish the same result:

Key Theorem 2.22 (adapted from [18, p. B1016])**.** *Let* $\mathbf{G}(s)$ *be the transfer function of a DAE-system with realization* $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$*. Let* $\mathbf{G}^{\mathrm{sp}}(s)$ *denote the strictly proper part and* $\mathbf{P}(s)$ *the improper part of* $\mathbf{G}(s)$*, i. e.* $\mathbf{G}(s) = \mathbf{G}^{\mathrm{sp}}(s) + \mathbf{P}(s)$*. Then the equalities*

$$\begin{aligned} \mathbf{G}^{\mathrm{sp}}(s) &= \mathbf{C}\,\boldsymbol{\Pi}_r^f\,(s\,\mathbf{E}-\mathbf{A})^{-1}\,\mathbf{B} = \mathbf{C}\,(s\,\mathbf{E}-\mathbf{A})^{-1}\,\boldsymbol{\Pi}_l^f\,\mathbf{B} \\ &= \mathbf{C}\,\boldsymbol{\Pi}_r^f\,(s\,\mathbf{E}-\mathbf{A})^{-1}\,\boldsymbol{\Pi}_l^f\,\mathbf{B}\,, \end{aligned} \tag{2.30}$$

and

$$\begin{aligned} \mathbf{P}(s) &= \mathbf{C}\,\boldsymbol{\Pi}_r^\infty\,(s\,\mathbf{E}-\mathbf{A})^{-1}\,\mathbf{B} = \mathbf{C}\,(s\,\mathbf{E}-\mathbf{A})^{-1}\,\boldsymbol{\Pi}_l^\infty\,\mathbf{B} \\ &= \mathbf{C}\,\boldsymbol{\Pi}_r^\infty\,(s\,\mathbf{E}-\mathbf{A})^{-1}\,\boldsymbol{\Pi}_l^\infty\,\mathbf{B} \end{aligned} \tag{2.31}$$

hold.

Proof. A reformulation of the first equality $\mathbf{G}^{\mathrm{sp}}(s) = \mathbf{C}\,\boldsymbol{\Pi}_r^f\,(s\,\mathbf{E}-\mathbf{A})^{-1}\,\mathbf{B}$ in Weierstraß canonical form leads to the definition of the strictly proper part according to (2.27):

$$\begin{aligned} \mathbf{G}^{\mathrm{sp}}(s) &= \mathbf{C}\,\mathbf{Q}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1}\left(s\,\mathbf{P}^{-1}\,\tilde{\mathbf{E}}\,\mathbf{Q}^{-1} - \mathbf{P}^{-1}\,\tilde{\mathbf{A}}\,\mathbf{Q}^{-1}\right)^{-1}\mathbf{P}^{-1}\,\tilde{\mathbf{B}} \\ &= \tilde{\mathbf{C}}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\left(s\,\tilde{\mathbf{E}} - \tilde{\mathbf{A}}\right)^{-1}\tilde{\mathbf{B}} = \tilde{\mathbf{C}}_f\left(s\,\mathbf{I}_{n_f} - \mathbf{J}\right)^{-1}\tilde{\mathbf{B}}_f\,. \end{aligned} \tag{2.32}$$

The remaining equalities can be verified in the same manner. ■

The partitioning of $\mathbf{G}(s)$ described in Theorem 2.22 allows to find detached realizations for the strictly proper and the improper subsystem:

Corollary 2.23. *Let $\mathbf{G}(s)$ be the transfer function of a DAE-system. If $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ is a realization of $\mathbf{G}(s)$, then*

- $\left[\mathbf{E},\, \mathbf{A},\, \mathbf{\Pi}_l^f\, \mathbf{B},\, \mathbf{C}\right]$, $\left[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}\, \mathbf{\Pi}_r^f\right]$ *and* $\left[\mathbf{E},\, \mathbf{A},\, \mathbf{\Pi}_l^f\, \mathbf{B},\, \mathbf{C}\, \mathbf{\Pi}_r^f\right]$ *are realizations of the strictly proper part $\mathbf{G}^{\mathrm{sp}}(s)$ and*
- $[\mathbf{E},\, \mathbf{A},\, \mathbf{\Pi}_l^\infty\, \mathbf{B},\, \mathbf{C}]$, $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}\, \mathbf{\Pi}_r^\infty]$ *and* $[\mathbf{E},\, \mathbf{A},\, \mathbf{\Pi}_l^\infty\, \mathbf{B},\, \mathbf{C}\, \mathbf{\Pi}_r^\infty]$ *are realizations of the improper part $\mathbf{P}(s)$.*

Proof. The result is a direct consequence of Theorem 2.22. ■

The statements of Theorem 2.22 and Corollary 2.23 allow two important conclusions: First, it does not matter if the left or right spectral projectors are used for partitioning. Therefore one has the choice to compute either a projected input matrix $\mathbf{B}$ or a projected output matrix $\mathbf{C}$ (or both). This circumstance will be exploited in Section 4.3 in order to reduce the computational effort. Second, the calculation of strictly proper and improper realizations keeps $\mathbf{E}$ and $\mathbf{A}$ unchanged. This is beneficial in a numerical point of view, since the sparsity is preserved.

2.2.3 Moments of a Transfer Function

The method of MOR discussed in this thesis is based on the interpolation of the transfer function in the frequency domain. For this purpose the *moments* of a transfer function are introduced:

Lemma 2.24 (adapted from [42, p. 17])**.** *Let $\mathbf{G}(s)$ be the transfer function of a DAE-system $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$. Then $\mathbf{G}(s)$ is given through*

$$\mathbf{G}(s) = \mathbf{C}\,(s\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{B} = -\sum_{\mu=0}^{\infty} \mathbf{M}^{(\mu)}(s_i)\,(s - s_i)^{\mu} \tag{2.33}$$

where $\mathbf{M}^{(\mu)}(s_i)$ denotes the μ-th moment *of $\mathbf{G}(s)$ around the point s_i and is defined as*

$$\mathbf{M}^{(\mu)}(s_i) = -\frac{1}{\mu!}\left(\frac{\mathrm{d}^{\mu}\mathbf{G}(s)}{\mathrm{d}s^{\mu}}\right)\Bigg|_{s=s_i} = \mathbf{C}\,\left[(\mathbf{A} - s_i\,\mathbf{E})^{-1}\,\mathbf{E}\right]^{\mu}\,(\mathbf{A} - s_i\,\mathbf{E})^{-1}\,\mathbf{B}\,, \tag{2.34}$$

for all $\mu \in \mathbb{N}^{\geq 0}$.

Proof. The result directly follows from the Taylor expansion of $\mathbf{G}(s)$ around the expansion point s_i. ■

Note that there are different conventions for the definition of the moments concerning the sign and the factor $\frac{1}{\mu!}$ (e. g. in comparison with [2, p. 345]). This way, no alternating signs are involved.

For the purpose of proving $\mathcal{H}_2$ pseudo-optimality in Chapter 4, a variation of Lemma 2.24 is given in Corollary 2.25.

Corollary 2.25. *Let $\mathbf{G}(s)$ be the transfer function of a strictly proper DAE-system with realization $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$. Then the μ-th moment of $\mathbf{G}(s)$ around the point s_i can be written as*

$$\mathbf{M}^{(\mu)}(s_i) = \mathbf{C} \left[(\mathbf{A} - s_i\, \mathbf{E})^{-1}\, \mathbf{E} \right]^{\mu} (\mathbf{A} - s_i\, \mathbf{E})^{-1}\, \mathbf{\Pi}_l^f\, \mathbf{B} \qquad \forall\, \mu \in \mathbb{N}^{\geq 0} . \tag{2.35}$$

Proof. According to Corollary 2.23 $[\mathbf{E},\, \mathbf{A},\, \mathbf{\Pi}_l^f\, \mathbf{B},\, \mathbf{C}]$ is a valid realization of $\mathbf{G}(s)$. Therefore the modification of (2.34) with $\mathbf{B} \rightarrow \mathbf{\Pi}_l^f\, \mathbf{B}$ does not change the transfer function or its moments. ■

2.2.4 Impulse Response of Strictly Proper DAEs

The impulse response $g(t) \in \mathbb{R}$ of a SISO system is defined as the output $y(t)$ corresponding to $\mathbf{x}(0) = \mathbf{0}$ and the special input $u(t) = \delta(t) \in \mathbb{R}$. In order to handle multiple-input, multiple-output (MIMO) systems a slightly enhanced relation is used: The impulse response $\mathbf{G}(t) \in \mathbb{R}^{p \times m}$ of a MIMO system is defined as the (combined) output corresponding to $\mathbf{x}(0) = \mathbf{0}$ and the special input $\mathbf{U}(t) = \delta(t)\, \mathbf{I}_m \in \mathbb{R}^{m \times m}$. Here the system input is no longer a vector $\mathbf{u}(t)$ but rather a matrix $\mathbf{U}(t)$. This way the entry in the v-th row and w-th column of $\mathbf{G}(t)$ describes the SISO impulse response between $u_w(t) = \delta(t)$ and $y_v(t)$.

Inserting $\mathbf{x}(0) = \mathbf{0}$ (and therefore $\tilde{\mathbf{x}}_{f0} = \mathbf{0}$) and $\mathbf{U}(t) = \delta(t)\, \mathbf{I}_m$ into (2.24) delivers with $t_0 = 0$ the impulse response of a strictly proper DAE:

$$\mathbf{G}(t) = \tilde{\mathbf{C}}_f \int_0^t e^{\mathbf{J}(t-\tau)}\, \tilde{\mathbf{B}}_f\, \delta(\tau)\, \mathrm{d}\tau . \tag{2.36}$$

This leads to following theorem:

Theorem 2.26. *Let $[\tilde{\mathbf{E}},\, \tilde{\mathbf{A}},\, \tilde{\mathbf{B}},\, \tilde{\mathbf{C}}]$ denote a realization of $\mathbf{G}(s)$ in Weierstraß canonical form according to (2.9). If the system is strictly proper, then its impulse response is given through*

$$\mathbf{G}(t) = \begin{cases} \mathbf{0} & , t < 0 \\ \tilde{\mathbf{C}}_f\, e^{\mathbf{J}\, t}\, \tilde{\mathbf{B}}_f & , t \geq 0 \end{cases} , \tag{2.37}$$

or equivalently

$$\mathbf{G}(t) = \begin{cases} \mathbf{0} & , t < 0 \\ \tilde{\mathbf{C}}\, e^{\tilde{\mathbf{E}}^{\mathrm{D}}\, \tilde{\mathbf{A}}\, t}\, \tilde{\mathbf{E}}^{\mathrm{D}}\, \tilde{\mathbf{B}} & , t \geq 0 \end{cases} . \tag{2.38}$$

Proof. The first equation is a direct consequence of (2.36) using the simplification

$$\int_{-\infty}^{\infty} f(t)\, \delta(t) \mathrm{d}t = f(0) . \tag{2.39}$$

In order to show (2.38), one has to compute the Drazin inverse of $\tilde{\mathbf{E}}$:

$$\tilde{\mathbf{E}}^{\mathrm{D}} = \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{N} \end{bmatrix}^{\mathrm{D}} \overset{\text{(Lemma 2.13)}}{=} \begin{bmatrix} \mathbf{I}_{n_f}^{\mathrm{D}} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}^{\mathrm{D}} \end{bmatrix} . \tag{2.40}$$

Using $\mathbf{I}_{n_f}^{\mathrm{D}} = \mathbf{I}_{n_f}$ (Lemma 2.11) and $\mathbf{N}^{\mathrm{D}} = \mathbf{0}$ (Lemma 2.12) leads to

$$\tilde{\mathbf{E}}^{\mathrm{D}} = \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} . \tag{2.41}$$

With the explicit knowledge of $\tilde{\mathbf{E}}^{\mathrm{D}}$ the equivalence of (2.37) and (2.38) can be shown:

$$\begin{aligned} \tilde{\mathbf{C}}\, e^{\tilde{\mathbf{E}}^{\mathrm{D}}\, \tilde{\mathbf{A}}\, t}\, \tilde{\mathbf{E}}^{\mathrm{D}}\, \tilde{\mathbf{B}} &= \tilde{\mathbf{C}} \left\{ \sum_{w=0}^{\infty} \left(\tilde{\mathbf{E}}^{\mathrm{D}}\, \tilde{\mathbf{A}} \right)^{w} \frac{t^w}{w!} \right\} \tilde{\mathbf{E}}^{\mathrm{D}}\, \tilde{\mathbf{B}} \\ &= \tilde{\mathbf{C}} \left\{ \sum_{w=0}^{\infty} \begin{bmatrix} \mathbf{J}^w & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \frac{t^w}{w!} \right\} \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{B}}_f \\ \tilde{\mathbf{B}}_\infty \end{bmatrix} \\ &= \begin{bmatrix} \tilde{\mathbf{C}}_f & \tilde{\mathbf{C}}_\infty \end{bmatrix} \begin{bmatrix} e^{\mathbf{J}\, t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{B}}_f \\ \mathbf{0} \end{bmatrix} = \tilde{\mathbf{C}}_f\, e^{\mathbf{J}\, t}\, \tilde{\mathbf{B}}_f \, . \end{aligned} \tag{2.42}$$

■

Remark 2.27. Since the impulse response $\mathbf{G}(t)$ is equivalent to the output $\mathbf{y}(t) \in \mathbb{R}^p$ related to a series of special inputs $\mathbf{u}(t) \in \mathbb{R}^m$ and all system matrices are assumed to be real-valued, $\mathbf{G}(t)$ is also real-valued: $\mathbf{G}(t) \in \mathbb{R}^{p \times m}$.

2.2.5 Controllability and Observability

In contrast to ODE-systems, there exist several controllability and observability concepts in the DAE-case [37, p. 35]. In the following only C- and R- controllability and observability are considered:

Definition 2.28 ([37, pp. 35,38])**.** The DAE-system $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ is called

- *completely controllable (C-controllable)*, if

$$\operatorname{rank} \begin{bmatrix} \varphi\, \mathbf{E} - \psi\, \mathbf{A} & \mathbf{B} \end{bmatrix} = n \qquad \forall\, (\varphi,\, \psi) \in \mathbb{C}^2 \setminus \{(0,\, 0)\} \, , \tag{2.43}$$

- *completely observable (C-observable)*, if

$$\operatorname{rank} \begin{bmatrix} \varphi\, \mathbf{E} - \psi\, \mathbf{A} \\ \mathbf{C} \end{bmatrix} = n \qquad \forall\, (\varphi,\, \psi) \in \mathbb{C}^2 \setminus \{(0,\, 0)\} \, , \tag{2.44}$$

- *controllable on a reachable set (R-controllable)*, if

$$\operatorname{rank} \begin{bmatrix} \lambda\, \mathbf{E} - \mathbf{A} & \mathbf{B} \end{bmatrix} = n \qquad \forall \text{ finite } \lambda \in \mathbb{C} \, , \tag{2.45}$$

- *observable on the reachable set (R-observable)*, if

$$\operatorname{rank} \begin{bmatrix} \lambda\, \mathbf{E} - \mathbf{A} \\ \mathbf{C} \end{bmatrix} = n \qquad \forall \text{ finite } \lambda \in \mathbb{C} \, . \tag{2.46}$$

Using the concept of controllability and observability, one can define the term “minimal realization” in the context of DAEs:

Definition 2.29 ([37, p. 8] and [29, p. 411])**.** A realization $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ of $\mathbf{G}(s)$ is called *minimal*, if the triplet $(\mathbf{E},\, \mathbf{A},\, \mathbf{B})$ is C-controllable and the triplet $(\mathbf{E},\, \mathbf{A},\, \mathbf{C})$ is C-observable. In this case, the dimension n of the matrices $\mathbf{E}$ and $\mathbf{A}$ is as small as possible.

This leads to following lemma:

Lemma 2.30. *If* $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ *is a realization of* $\mathbf{G}(s)$ *with* $\det(\mathbf{E}) = 0$ *and the DAE-system is strictly proper, then the realization is not minimal.*

Proof. Due to singular $\mathbf{E}$, there exists a fast subsystem $\tilde{\mathbf{x}}_\infty(t)$, i. e. $n_\infty > 0$ with $n_f + n_\infty = n$. Since the transfer function is strictly proper the fast subsystem does not influence the output $\mathbf{y}(t)$. Therefore a smaller realization $[\hat{\mathbf{E}},\, \hat{\mathbf{A}},\, \hat{\mathbf{B}},\, \hat{\mathbf{C}}]$ of $\mathbf{G}(s)$ with $\det(\hat{\mathbf{E}}) \neq 0$ and $\hat{n} = n_f < n$ (i. e. an ODE-system) exists. Hence the realization $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ is not minimal according to Definition 2.29. ■

Since properness is strongly related to the causality of a system, many technical applications deal with strictly proper DAEs. Therefore one must assume that a FOM provided as a DAE-system might be not minimal. However, this can be considered as an advantage, since in this case it is possible to find a suitable ROM in ODE-form (which is beneficial especially if the ROM is intended to be used for simulation):

Corollary 2.31. *For every strictly proper DAE-system there exists a realization in ODE-form, i. e.* $\det(\mathbf{E}) \neq 0$.

Proof. The proof is contained in Lemma 2.30. ■

Note that in technical applications improper DAEs occur as well: consider a single mass excited by a predefined force (input). If the position of the mass is defined as output, the transfer function is strictly proper. In contrast an improper system arises, if the jerk (derivative of acceleration with respect to time) is chosen as output. To summarize, the matrices $\mathbf{E}$ and $\mathbf{A}$ determine the index of a DAE-system, while the choice of inputs and outputs (through $\mathbf{B}$ and $\mathbf{C}$) additionally affects the properness.

Apart from spectral projectors, the controllability and observability *Gramians* are of great importance. They play a key role in MOR by BT, but are also connected to Krylov-based methods as Chapter 4 will show. In the DAE-case one distinguishes between proper and improper Gramians as discussed in [37] in detail (originally defined in [8]):

Definition 2.32 (adapted from [29, p. 412])**.** Let $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ denote a realization of an asymptotically stable DAE-system. Then the *proper controllability* and *observability Gramians* $\mathbf{\Gamma}^{\mathrm{pc}}$ and $\mathbf{\Gamma}^{\mathrm{po}}$ are defined as the unique Hermitian, positive semidefinite solutions of the *generalized projected continuous-time Lyapunov equations*

$$\mathbf{A}\,\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{E}^* + \mathbf{E}\,\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{A}^* + \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{B}^*\,\mathbf{\Pi}_l^{f*} = \mathbf{0}\,, \qquad \mathbf{\Gamma}^{\mathrm{pc}} = \mathbf{\Pi}_r^f\,\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{\Pi}_r^{f*}\,, \tag{2.47}$$

$$\mathbf{A}^*\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{E} + \mathbf{E}^*\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{A} + \mathbf{\Pi}_r^{f*}\,\mathbf{C}^*\,\mathbf{C}\,\mathbf{\Pi}_r^f = \mathbf{0}\,, \qquad \mathbf{\Gamma}^{\mathrm{po}} = \mathbf{\Pi}_l^{f*}\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{\Pi}_l^f\,. \tag{2.48}$$

Further the *improper controllability* and *observability Gramians* $\mathbf{\Gamma}^{\mathrm{imc}}$ and $\mathbf{\Gamma}^{\mathrm{imo}}$ are defined as the unique Hermitian, positive semidefinite solutions of the *generalized projected discrete-time Lyapunov equations*

$$\mathbf{A}\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{A}^* - \mathbf{E}\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{E}^* - \mathbf{\Pi}_l^\infty\,\mathbf{B}\,\mathbf{B}^*\,\mathbf{\Pi}_l^{\infty*} = \mathbf{0}\,, \qquad \mathbf{\Gamma}^{\mathrm{imc}} = \mathbf{\Pi}_r^\infty\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{\Pi}_r^{\infty*}\,, \tag{2.49}$$

$$\mathbf{A}^*\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{A} - \mathbf{E}^*\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{E} - \mathbf{\Pi}_r^{\infty*}\,\mathbf{C}^*\,\mathbf{C}\,\mathbf{\Pi}_r^\infty = \mathbf{0}\,, \qquad \mathbf{\Gamma}^{\mathrm{imo}} = \mathbf{\Pi}_l^{\infty*}\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{\Pi}_l^\infty\,. \tag{2.50}$$

Chapter 3

Model Order Reduction via Tangential Interpolation

After the introduction into linear DAE-theory, the basic reduction scheme is explained in the following sections. First the concept of tangential interpolation of MIMO systems known from ODE-MOR (see [15]) is presented. Subsequently rational Krylov subspaces and generalized Sylvester equations are introduced, which can be used to describe the interpolation process and act as main tools in the following chapters. Finally the general framework for reducing improper DAE-systems (which is based on a separate reduction of the strictly proper and improper subsystem in Chapter 4 and Chapter 5) is explained.

3.1 Problem Statement

In this section the basic idea of moment matching for SISO-systems and especially the extension to MIMO-systems alias MOR by tangential interpolation are presented. These concepts can be considered as subgoals (beside $\mathcal{H}_2$ pseudo-optimality) of the MOR-techniques discussed in this contribution. This section gives a brief overview, while the actual algorithms which enforce moment matching/tangential interpolation are derived in Chapter 4.

MOR by moment matching is based on the approximation of the transfer function in the frequency domain. More precisely the transfer function is interpolated at specific expansion points (known as *shifts*) using the coefficients (called *moments*) defined in Lemma 2.24. Depending on the order of the moments, the transfer function and its derivatives (with respect to s) are matched. A mathematical formulation of moment matching in the SISO-case reads as follows:

Definition 3.1 (adapted from [2, p. 345f.]). Let $\mathbf{G}(s)$, $\mathbf{G}_\mathrm{r}(s) \in \mathbb{C}^{1\times 1}$ denote the transfer functions of the SISO-type FOM and ROM respectively. Then the ROM matches the first $\rho_i \in \mathbb{N}^{>0}$ moments of the FOM at the expansion point s_i, if

$$\left(\frac{\mathrm{d}^\mu \mathbf{G}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} = \left(\frac{\mathrm{d}^\mu \mathbf{G}_\mathrm{r}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \qquad \forall\, \mu = 0, \dots, \rho_i - 1 \tag{3.1}$$

holds.

Obviously a high number of matched moments ρ_i leads in general to a good approximation in the surrounding of s_i. Furthermore note that Definition 3.1 does not distinguish between ODE- and DAE-systems at all.

As an example, Figure 3.1 illustrates the approximation of an improper index 2 SISO-DAE. Therein a reduction by (ODE-) CUREd SPARK (see [30] and [42]) was used, i. e. the special characteristics of DAEs have not been addressed. A rather low dimension of the ROM ($q = 2$) was chosen in order to show small error in the surrounding of the expansion points $s_{1,2} = -1 \pm 10\imath$ (corresponds to the peak at $\omega = 10\,\mathrm{rad/s}$), while the remaining spectrum is poorly approximated.

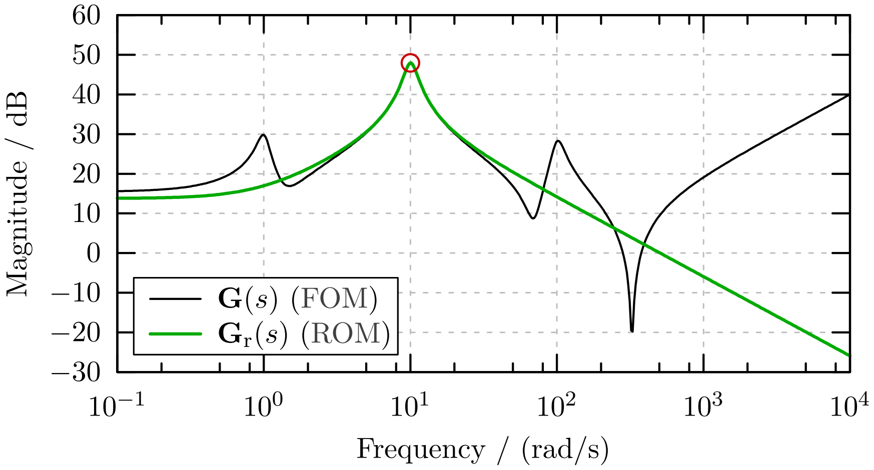


Figure 3.1: Moment matching of an improper index 2 SISO-DAE by (ODE-) CUREd SPARK: the transfer function of the FOM is approximated by a ROM of dimension $q = 2$. The red circle indicates matching of the first moment ($\rho_1 = \rho_2 = 1$) at the expansion points $s_{1,2} = -1 \pm 10\imath$.

Figure 3.1 demonstrates moment matching in the SISO-case. To reduce MIMO systems one has to distinguish between different input-output combinations (called *channels*) which correspond to the different matrix entries of $\mathbf{G}(s) \in \mathbb{C}^{p\times m}$. In order to interpolate channels (or combinations of them) at different expansion points, *tangential interpolation* is used: as an extension to moment matching of SISO-systems (Definition 3.1), the *tangential directions* $\mathbf{r}_{ij}$ (right) and $\mathbf{l}_{ij}$ (left) are introduced, which describe a weighted approximation:

Definition 3.2 (adapted from [15, p. 329f.])**.** Let $\mathbf{G}(s)$, $\mathbf{G}_\mathrm{r}(s) \in \mathbb{C}^{p\times m}$ denote the transfer functions of the MIMO-type FOM and ROM respectively. Then

- *right tangential interpolation* at the expansion point s_i, in direction $\mathbf{r}_{ij} \in \mathbb{C}^{m\times 1}\backslash\{\mathbf{0}\}$ and of order $\rho_{ij} \in \mathbb{N}^{>0}$ is defined as

$$\left(\frac{\mathrm{d}^\mu \mathbf{G}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} = \left(\frac{\mathrm{d}^\mu \mathbf{G}_\mathrm{r}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} \qquad \forall\, \mu = 0, \ldots, \rho_{ij} - 1\,, \tag{3.2}$$

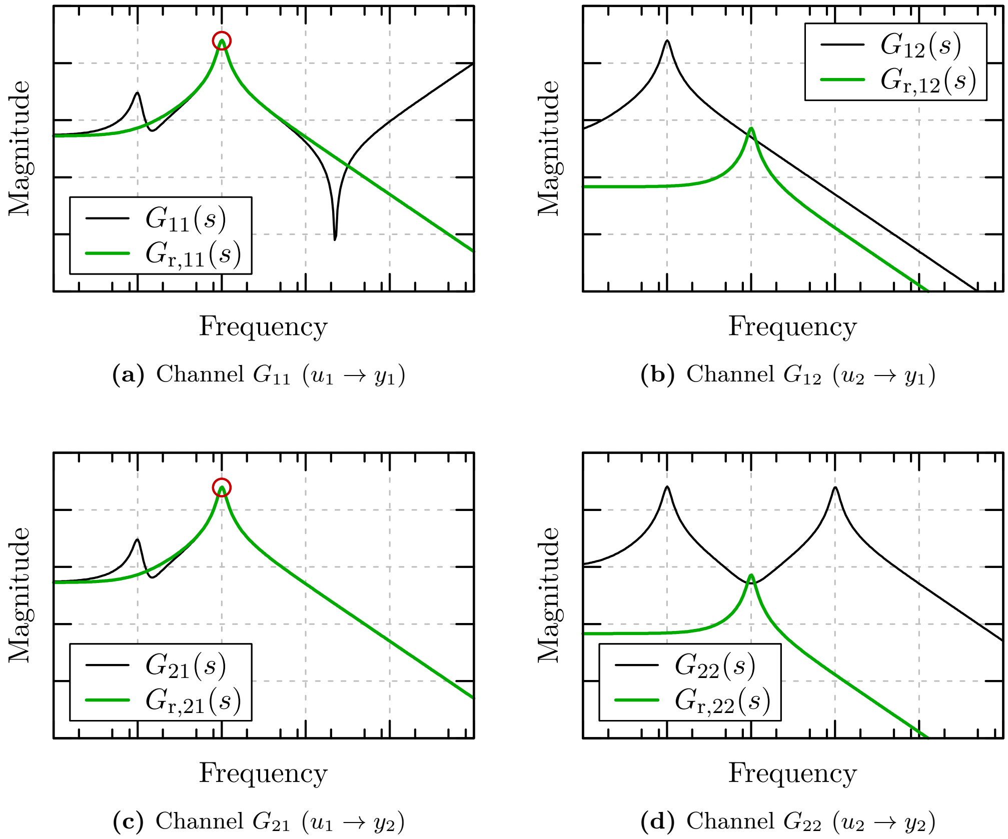
- *left tangential interpolation* at the expansion point s_i, in direction $\mathbf{l}_{ij} \in \mathbb{C}^{1\times p} \setminus \{\mathbf{0}\}$ and of order $\rho_{ij} \in \mathbb{N}^{>0}$ is defined as

$$\mathbf{l}_{ij}^\mathrm{T} \cdot \left(\frac{\mathrm{d}^\mu \mathbf{G}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} = \mathbf{l}_{ij}^\mathrm{T} \cdot \left(\frac{\mathrm{d}^\mu \mathbf{G}_\mathrm{r}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \qquad \forall\, \mu = 0, \ldots, \rho_{ij} - 1\,, \tag{3.3}$$

- *two-sided tangential interpolation* at the expansion point s_i, in the directions $\mathbf{r}_{ij} \in \mathbb{C}^{m\times 1} \setminus \{\mathbf{0}\}$ and $\mathbf{l}_{ij} \in \mathbb{C}^{1\times p} \setminus \{\mathbf{0}\}$ and of order $\rho_{ij} \in \mathbb{N}^{>0}$ is defined as

$$\mathbf{l}_{ij}^{\mathrm{T}} \cdot \left(\frac{\mathrm{d}^{\mu}\mathbf{G}(s)}{\mathrm{d}s^{\mu}}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} = \mathbf{l}_{ij}^{\mathrm{T}} \cdot \left(\frac{\mathrm{d}^{\mu}\mathbf{G}_{\mathrm{r}}(s)}{\mathrm{d}s^{\mu}}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} \qquad \forall\, \mu = 0, \dots, \rho_{ij} - 1 \,. \quad (3.4)$$

Note that in contrast to SISO-moment matching, one has additionally to specify tangential directions, which on the one hand increases the degrees of freedom during reduction. On the other hand the proper choice of $\mathbf{r}_{ij}$ and $\mathbf{l}_{ij}$ in order to obtain good approximation results seems to be a difficult task. This is demonstrated in Figure 3.2, which compares the original and reduced frequency responses in different channels of an (improper) MIMO-DAE-system.



(a) Channel G_{11} $(u_1 \to y_1)$

(b) Channel G_{12} $(u_2 \to y_1)$

(c) Channel G_{21} $(u_1 \to y_2)$

(d) Channel G_{22} $(u_2 \to y_2)$

Figure 3.2: (Right) tangential interpolation of an improper MIMO DAE-system. The ROM is obtained by (ODE-) input PORK (see [42]). While the transfer functions of the FOM and the ROM are matched at a special point (red circle) in (a) and (c) (channel G_{11} and G_{21}), no moment matching is achieved in (b) and (d) (channel G_{12} and G_{22}).

Since the dimension of the transfer function is chosen to 2 × 2, i. e. $p = 2$ outputs and $m = 2$ inputs, four distinct channels exist:

$$\mathbf{G}(s) = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \in \mathbb{C}^{2\times 2} \quad \text{and} \quad \mathbf{G}_{\mathrm{r}}(s) = \begin{bmatrix} G_{\mathrm{r},11}(s) & G_{\mathrm{r},12}(s) \\ G_{\mathrm{r},21}(s) & G_{\mathrm{r},22}(s) \end{bmatrix} \in \mathbb{C}^{2\times 2} \,, \quad (3.5)$$

where $G_{vw}(s)$ and $G_{\mathrm{r},vw}(s)$ denote the input-output behavior of the FOM and ROM corresponding to the w-th input $u_w(t)$ and v-th output $y_v(t)$.

As depicted in Figure 3.2, moment matching achieved in one channel does in general not apply to the remaining ones. This way a selective (or weighted) approximation is possible. If all channels are of interest, one may enforce matching at the peaks of the frequency response for each channel separately to obtain a reasonable ROM.

Remark 3.3. Within the scope of this thesis it is assumed, that expansion points are chosen in the open right half of the complex plane. This is essential for the proof of $\mathcal{H}_2$ pseudo-optimality in Chapter 4. As a consequence one has to pay special attention in order to avoid stationary errors (i. e. deviations of the ROM from the FOM at $s = 0$).

Remark 3.4. During adaptation of MOR techniques to the DAE-case one may think of using *Markov parameters* (additional to moments) for interpolation. Those are used to match the transfer function at infinity, i. e.

$$\lim_{s \to \infty} \left(\frac{\mathrm{d}^{\mu} \mathbf{G}(s)}{\mathrm{d}s^{\mu}} \right) = \lim_{s \to \infty} \left(\frac{\mathrm{d}^{\mu} \mathbf{G}_{\mathrm{r}}(s)}{\mathrm{d}s^{\mu}} \right) . \tag{3.6}$$

Keep in mind, that this originates from the ODE-case and is <u>not</u> equivalent to matching the polynomial part $\mathbf{P}(s)$ of a DAE, which has to be handled separately. Because the contribution of $\mathbf{P}(s)$ dominates at high frequencies (and thus at $\omega \to \infty$), Markov parameters will not be considered.

3.2 Rational Krylov Subspaces

In the following rational Krylov subspaces are introduced, which will be used in Section 3.5 to achieve tangential interpolation during projective MOR. Aside from that, they are the basis of the investigated $\mathcal{H}_2$ pseudo-optimal reduction scheme.

In mathematics Krylov subspaces are spanned by a vector and its multiplication with a predefined matrix:

Definition 3.5 ([2, p. 313])**.** Let $\mathbf{X} \in \mathbb{C}^{u \times u}$, $\mathbf{y} \in \mathbb{C}^{u \times 1}$ and $w \in \mathbb{N}^{>0}$. Then the space

$$\mathcal{K}_w(\mathbf{X},\, \mathbf{y}) := \operatorname{span}\left\{ \mathbf{y},\, \mathbf{X}\,\mathbf{y},\, \mathbf{X}^2\,\mathbf{y},\, \dots,\, \mathbf{X}^{(w-1)}\,\mathbf{y} \right\} \tag{3.7}$$

is called *Krylov subspace of order w*.

In linear system theory one often has to deal with the expression $(s\,\mathbf{E} - \mathbf{A})^{-1}$ since it is part of *rational* transfer functions and their moments. In preparation for moment matching *rational* Krylov subspaces are defined:

Definition 3.6. Let the FOM be described by the DAE-system $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ with $\mathbf{E},\, \mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$ and $\mathbf{C} \in \mathbb{R}^{p \times n}$. Then

- the space

$$\mathcal{K}^{i}_{\mathrm{bi}}\,(s_i,\, q_i) := \mathcal{K}_{q_i}\left((\mathbf{A} - s_i\,\mathbf{E})^{-1}\,\mathbf{E},\, (\mathbf{A} - s_i\,\mathbf{E})^{-1}\,\mathbf{B} \right) , \tag{3.8}$$

is called *block-input rational Krylov subspace* of the FOM at the expansion point $s_i \in \mathbb{C}$ and of order $q_i \in \mathbb{N}^{>0}$, and

- the space

$$\mathcal{K}_{\text{bo}}^{i}\left(s_i,\, q_i\right) := \mathcal{K}_{q_i}\left(\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-\text{T}}\mathbf{E}^{\text{T}},\, \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-\text{T}}\mathbf{C}^{\text{T}}\right)\ , \tag{3.9}$$

  is called *block-output rational Krylov subspace* of the FOM at the expansion point $s_i \in \mathbb{C}$ and of order $q_i \in \mathbb{N}^{>0}$, and

- the space

$$\mathcal{K}_{\text{ti}}^{ij}\left(s_i,\, \mathbf{r}_{ij},\, q_{ij}\right) := \mathcal{K}_{q_{ij}}\left(\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{E},\, \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{B}\,\mathbf{r}_{ij}\right)\ , \tag{3.10}$$

  is called *tangential-input rational Krylov subspace* of the FOM at the expansion point $s_i \in \mathbb{C}$, in tangential direction $\mathbf{r}_{ij} \in \mathbb{C}^{m\times 1}$ and of order $q_{ij} \in \mathbb{N}^{>0}$, and

- the space

$$\mathcal{K}_{\text{to}}^{ij}\left(s_i,\, \mathbf{l}_{ij},\, q_{ij}\right) := \mathcal{K}_{q_{ij}}\left(\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-\text{T}}\mathbf{E}^{\text{T}},\, \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-\text{T}}\mathbf{C}^{\text{T}}\,\mathbf{l}_{ij}\right)\ , \tag{3.11}$$

  is called *tangential-output rational Krylov subspace* of the FOM at the expansion point $s_i \in \mathbb{C}$, in tangential direction $\mathbf{l}_{ij} \in \mathbb{C}^{p\times 1}$ and of order $q_{ij} \in \mathbb{N}^{>0}$.

While block Krylov subspaces treat all channels of the transfer function in the same way, different interpolation data can be defined with tangential Krylov subspaces (tangential interpolation). Therefore $\mathcal{K}_{\text{ti}}$ and $\mathcal{K}_{\text{to}}$ can be considered as a generalization of $\mathcal{K}_{\text{bi}}$ and $\mathcal{K}_{\text{bo}}$. In fact, every block-input and block-output rational Krylov subspace can be written as a tangential-input and tangential-output rational Krylov subspace respectively:

Corollary 3.7. *Let all assumptions of Definition 3.6 hold and let $\mathbf{e}_j$ denote the j-th unit vector in $\mathbb{R}^m$, i. e.*

$$\mathbf{e}_j = \left[e_{j,1},\, \dots,\, e_{j,u},\, \dots,\, e_{j,m}\right]^{\text{T}} \in \mathbb{R}^{m\times 1} \quad \textit{with} \quad e_{j,u} = \begin{cases} 1 & \textit{for } u = j \\ 0 & \textit{for } u \neq j \end{cases}\ . \tag{3.12}$$

Then

$$\mathcal{K}_{\text{bi}}^{i}\left(s_i,\, q_i\right) = \bigcup_{j=1}^{m} \mathcal{K}_{\text{ti}}^{ij}\left(s_i,\, \mathbf{e}_j,\, q_i\right) \quad \textit{and} \quad \mathcal{K}_{\text{bo}}^{i}\left(s_i,\, q_i\right) = \bigcup_{j=1}^{p} \mathcal{K}_{\text{to}}^{ij}\left(s_i,\, \mathbf{e}_j,\, q_i\right) \tag{3.13}$$

holds.

Proof. Note that the use of m unit vectors $\mathbf{e}_1, \dots, \mathbf{e}_m$ as tangential directions $\mathbf{r}_{ij}$ is equivalent to a multiplication of $\mathbf{B}$ with the identity $\mathbf{I}_m$ from the right. The same holds for block-output rational Krylov subspaces. ■

To keep following proofs as general as possible, only tangential-input and tangential-output rational Krylov subspaces will be discussed.

The construction of rational Krylov subspaces according to Definition 3.6 leads to sequences of vectors getting closer and closer to linear dependence with increasing w. This causes great numerical error in the following reduction process. Therefore one tries to find a different set of vectors which describe the same subspace:

Definition 3.8. Let $\mathbf{Z}$ be a matrix of full column rank which fulfils

$$\mathcal{R}(\mathbf{Z}) = \operatorname{colspan}(\mathbf{Z}) = \mathcal{K}_w(\mathbf{X},\, \mathbf{y}) \,. \tag{3.14}$$

Then $\mathbf{Z}$ is called *basis* of the Krylov subspace $\mathcal{K}_w(\mathbf{X},\, \mathbf{y})$.

Corollary 3.9. *Let* $\mathbf{Z}$ *be a basis of* $\mathcal{K}_w(\mathbf{X},\, \mathbf{y})$ *as introduced in Definition 3.8. Then* $\hat{\mathbf{Z}} = \mathbf{Z}\,\mathbf{T}$*, whereat* $\mathbf{T}$ *is a quadratic and regular matrix of appropriate dimension, is also a basis of* $\mathcal{K}_w(\mathbf{X},\, \mathbf{y})$*, i. e.* $\mathcal{R}(\mathbf{Z}) = \mathcal{R}(\hat{\mathbf{Z}}) = \mathcal{K}_w(\mathbf{X},\, \mathbf{y})$*.*

Ideally the basis of a Krylov subspace is orthonormal, such that subsequent algorithms do not suffer from bad numerical condition. To achieve that, several orthogonalization techniques like the Gram-Schmidt process (as part of the rational Arnoldi algorithm [2, p. 335f.]) are available. Nevertheless the result of the construction scheme described in Definition 3.6 is used for proofs:

Definition 3.10. Let $\mathbf{Z}^{\mathrm{P}}$ be a matrix of full column rank which is constructed as

$$\mathbf{Z}^{\mathrm{P}} = \left[\mathbf{y},\, \mathbf{X}\,\mathbf{y},\, \mathbf{X}^2\,\mathbf{y},\, \ldots,\, \mathbf{X}^{(w-1)}\,\mathbf{y}\right] \,. \tag{3.15}$$

Then $\mathbf{Z}^{\mathrm{P}}$ is called *primitive* basis of the Krylov subspace $\mathcal{K}_w(\mathbf{X},\, \mathbf{y})$. Note the superscript P, which indicates, that the basis is *primitive*.

Remark 3.11. Although Krylov subspaces technically have a structure as defined in (3.7), also unions $\mathcal{K}_{1\cup 2\cup 3\cup\ldots} = \mathcal{K}_{w_1}(\mathbf{X}_1,\, \mathbf{y}_1) \cup \mathcal{K}_{w_2}(\mathbf{X}_2,\, \mathbf{y}_2) \cup \mathcal{K}_{w_3}(\mathbf{X}_3,\, \mathbf{y}_3) \cup \ldots$ will be called *accumulated* Krylov subspaces.

Remark 3.12. The vectors $\{\mathbf{y},\, \mathbf{X}\,\mathbf{y},\, \mathbf{X}^2\,\mathbf{y},\, \ldots\}$ used in (3.7) to define the Krylov subspace $\mathcal{K}_w(\mathbf{X},\, \mathbf{y})$ may not be linearly independent. This is obviously the case for $w > u$ (with $\mathbf{X} \in \mathbb{C}^{u\times u}$) but may also apply for $w < u$. Furthermore the concatenation of the bases of two Krylov subspaces may contain linearly dependent vectors.

In order to get a basis of the Krylov subspace, redundant directions have to be truncated, which is not subject of this work. Therefore it is assumed in the following, that the vectors $\{\mathbf{y},\, \mathbf{X}\,\mathbf{y},\, \mathbf{X}^2\,\mathbf{y},\, \ldots\}$ used during the construction are linearly independent, i. e. they describe the primitive base of the Krylov subspace.

Accumulation

This section is based on [42, p. 27ff.] and aims to reveal how the construction of accumulated rational Krylov subspaces can be described by matrix equations. In the following, a union of tangential-input rational Krylov subspaces according to Definition 3.6 is built up, which is basically done by stacking their primitive bases. It is important to note that Definition 3.6 differs from the notations in [42]. However both representations can be used to describe the same Krylov subspace. Due to the duality principle in linear systems (see [2, p. 76]), tangential-output rational Krylov subspaces are not treated, since all proofs work in a dual way.

The accumulation scheme used in the following is illustrated in Figure 3.3. To interpolate the FOM at different frequencies, s different expansion points $\{s_1,\, \ldots,\, s_i,\, \ldots,\, s_s\}$ are used. For each expansion point s_i, a set of r_i tangential directions $\{\mathbf{r}_{i1},\, \ldots,\, \mathbf{r}_{ij},\, \ldots,\, \mathbf{r}_{ir_i}\}$ is defined, which encodes the weights of the channels of $\mathbf{G}(s)$ used for moment matching.

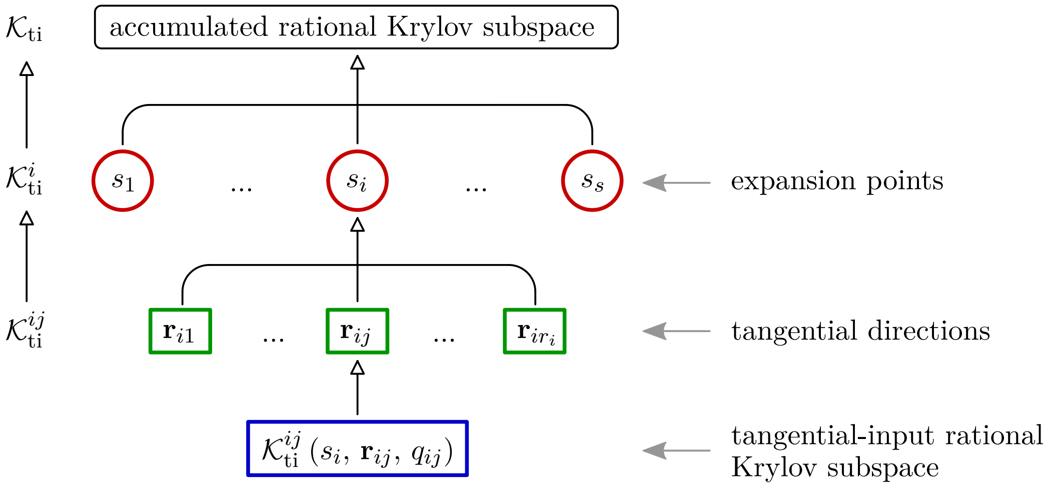

$\mathcal{K}_{\text{ti}}$
accumulated rational Krylov subspace
$\mathcal{K}_{\text{ti}}^{i}$
s_1
...
s_i
...
s_s
expansion points
$\mathcal{K}_{\text{ti}}^{ij}$
$\mathbf{r}_{i1}$
...
$\mathbf{r}_{ij}$
...
$\mathbf{r}_{ir_i}$
tangential directions
$\mathcal{K}_{\text{ti}}^{ij}\left(s_i,\, \mathbf{r}_{ij},\, q_{ij}\right)$
tangential-input rational Krylov subspace

Figure 3.3: Accumulation scheme of rational Krylov subspaces. The tangential-input rational Krylov subspaces $\mathcal{K}_{\text{ti}}^{ij}$ related to the tangential directions $\mathbf{r}_{ij}$ of an expansion point s_i are combined to $\mathcal{K}_{\text{ti}}^{i}$. These in turn are concatenated to the accumulated rational Krylov subspace $\mathcal{K}_{\text{ti}}$.

Moreover each tangential direction $\mathbf{r}_{ij}$ is associated with a rational Krylov subspace $\mathcal{K}_{\text{ti}}^{ij}$ of order q_{ij}, where q_{ij} determines the level of interpolation (highest derivatives).

According to Definition 3.6 the tangential-input rational Krylov subspace at expansion point $s_i \in \mathbb{C}$, in tangential direction $\mathbf{r}_{ij} \in \mathbb{C}^{m\times 1}$ and of order $q_{ij} \in \mathbb{N}^{>0}$ is given through

$$\begin{aligned}\mathcal{K}_{\text{ti}}^{ij}\left(s_i,\, \mathbf{r}_{ij},\, q_{ij}\right) := \operatorname{span}\Big\{ & \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{B}\,\mathbf{r}_{ij},\\ & \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{E}\,\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{B}\,\mathbf{r}_{ij},\\ & \qquad \vdots \\ & \left[\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{E}\right]^{q_{ij}-1}\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{B}\,\mathbf{r}_{ij}\Big\}\,.\end{aligned} \tag{3.16}$$

Using the auxiliary vectors

$$\begin{aligned}\mathbf{v}_{ij1}^{\text{P}} &:= \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{B}\,\mathbf{r}_{ij}\ ,\\ \mathbf{v}_{ij2}^{\text{P}} &:= \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{E}\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{B}\,\mathbf{r}_{ij} &&= \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{E}\,\mathbf{v}_{ij1}^{\text{P}}\ ,\\ &\ \,\vdots\\ \mathbf{v}_{ijq_{ij}}^{\text{P}} &:= \left[\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{E}\right]^{q_{ij}-1}\left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{B}\,\mathbf{r}_{ij} &&= \left(\mathbf{A} - s_i\,\mathbf{E}\right)^{-1}\mathbf{E}\,\mathbf{v}_{ij(q_{ij}-1)}^{\text{P}}\end{aligned} \tag{3.17}$$

leads to the shorter form

$$\mathcal{K}_{\text{ti}}^{ij} = \operatorname{span}\left\{\mathbf{v}_{ij1}^{\text{P}},\, \dots,\, \mathbf{v}_{ijq_{ij}}^{\text{P}}\right\}\ . \tag{3.18}$$

After introducing the matrix

$$\mathbf{V}_{ij}^{\text{P}} := \left[\mathbf{v}_{ij1}^{\text{P}},\, \dots,\, \mathbf{v}_{ijq_{ij}}^{\text{P}}\right] \in \mathbb{C}^{n\times q_{ij}} \tag{3.19}$$

as the primitive basis of $\mathcal{K}_{\text{ti}}^{ij}$ one can describe $\mathcal{K}_{\text{ti}}^{ij}$ as the image of $\mathbf{V}_{ij}^{\text{P}}$, i.e. $\mathcal{K}_{\text{ti}}^{ij} = \mathcal{R}\left\{\mathbf{V}_{ij}^{\text{P}}\right\}$. Rearranging (3.17) to

$$\begin{aligned}(\mathbf{A} - s_i\,\mathbf{E})\,\mathbf{v}_{ij1}^{\text{P}} &= \mathbf{B}\,\mathbf{r}_{ij}\\ (\mathbf{A} - s_i\,\mathbf{E})\,\mathbf{v}_{ij2}^{\text{P}} &= \mathbf{E}\,\mathbf{v}_{ij1}^{\text{P}}\\ &\vdots\\ (\mathbf{A} - s_i\,\mathbf{E})\,\mathbf{v}_{ijq_{ij}}^{\text{P}} &= \mathbf{E}\,\mathbf{v}_{ij(q_{ij}-1)}^{\text{P}}\end{aligned} \tag{3.20}$$

helps assembling the matrix equation

$$\mathbf{A}\underbrace{\left[\mathbf{v}_{ij1}^{\text{P}},\,\dots,\,\mathbf{v}_{ijq_{ij}}^{\text{P}}\right]}_{\mathbf{V}_{ij}^{\text{P}}} - \mathbf{E}\underbrace{\left[\mathbf{v}_{ij1}^{\text{P}},\,\dots,\,\mathbf{v}_{ijq_{ij}}^{\text{P}}\right]}_{\mathbf{V}_{ij}^{\text{P}}}\underbrace{\begin{bmatrix} s_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & s_i\end{bmatrix}}_{\mathbf{S}_{V,ij}^{\text{P}}} = \mathbf{B}\underbrace{\left[\mathbf{r}_{ij},\,\mathbf{0},\,\dots,\,\mathbf{0}\right]}_{\mathbf{R}_{ij}^{\text{P}}} \tag{3.21}$$

or in the more compact form

$$\mathbf{A}\,\mathbf{V}_{ij}^{\text{P}} - \mathbf{E}\,\mathbf{V}_{ij}^{\text{P}}\,\mathbf{S}_{V,ij}^{\text{P}} = \mathbf{B}\,\mathbf{R}_{ij}^{\text{P}}\ , \tag{3.22}$$

whereat the Jordan-block $\mathbf{S}_{V,ij}^{\text{P}} \in \mathbb{C}^{q_{ij}\times q_{ij}}$ contains the information about the expansion points and the tangential direction is considered through $\mathbf{R}_{ij}^{\text{P}} \in \mathbb{C}^{m\times q_{ij}}$.

In order to handle r_i tangential directions $\mathbf{r}_{ij}$ at the expansion point s_i, i.e. $j \in \{1,\,2,\,\dots,\,r_i\}$, a new matrix equation is assembled (out of r_i-equations (3.22) for each tangential direction):

$$\mathbf{A}\,\mathbf{V}_{i}^{\text{P}} - \mathbf{E}\,\mathbf{V}_{i}^{\text{P}}\,\mathbf{S}_{V,i}^{\text{P}} = \mathbf{B}\,\mathbf{R}_{i}^{\text{P}} \tag{3.23}$$

with the new matrices

$$\begin{aligned}\mathbf{V}_{i}^{\text{P}} &= \left[\mathbf{V}_{i1}^{\text{P}},\,\dots,\,\mathbf{V}_{ir_i}^{\text{P}}\right] \in \mathbb{C}^{n\times q_i}\ ,\\ \mathbf{S}_{V,i}^{\text{P}} &= \operatorname{diag}\left(\mathbf{S}_{V,i1}^{\text{P}},\,\dots,\,\mathbf{S}_{V,ir_i}^{\text{P}}\right) \in \mathbb{C}^{q_i\times q_i}\ ,\\ \mathbf{R}_{i}^{\text{P}} &= \left[\mathbf{R}_{i1}^{\text{P}},\,\dots,\,\mathbf{R}_{ir_i}^{\text{P}}\right] \in \mathbb{C}^{m\times q_i}\end{aligned} \tag{3.24}$$

and $q_i = \sum_{j=1}^{r_i} q_{ij}$. This leads to the expanded Krylov subspace

$$\mathcal{K}_{\text{ti}}^{i} = \bigcup_{j=1}^{r_i} \mathcal{K}_{\text{ti}}^{ij} = \mathcal{R}\left\{\mathbf{V}_{i}^{\text{P}}\right\}\ . \tag{3.25}$$

In the same way s different expansion points, i.e. $i \in \{1,\,2,\,\dots,\,s\}$, can be handled, which results in

$$\mathbf{A}\,\mathbf{V}^{\text{P}} - \mathbf{E}\,\mathbf{V}^{\text{P}}\,\mathbf{S}_{V}^{\text{P}} = \mathbf{B}\,\mathbf{R}^{\text{P}}\ , \tag{3.26}$$

with

$$\begin{aligned}\mathbf{V}^{\text{P}} &= \left[\mathbf{V}_{1}^{\text{P}},\,\dots,\,\mathbf{V}_{s}^{\text{P}}\right] \in \mathbb{C}^{n\times q}\ ,\\ \mathbf{S}_{V}^{\text{P}} &= \operatorname{diag}\left(\mathbf{S}_{V,1}^{\text{P}},\,\dots,\,\mathbf{S}_{V,s}^{\text{P}}\right) \in \mathbb{C}^{q\times q} \quad \text{(in Jordan canonical form)}\ ,\\ \mathbf{R}^{\text{P}} &= \left[\mathbf{R}_{1}^{\text{P}},\,\dots,\,\mathbf{R}_{s}^{\text{P}}\right] \in \mathbb{C}^{m\times q}\end{aligned} \tag{3.27}$$

and $q = \sum_{i=1}^{s} q_i$. Matrix equations with the shape of (3.26) are called *generalized Sylvester equations*. The *accumulated* tangential-input rational Krylov subspace for multiple expansion points (with each containing individual tangential directions) finally reads as

$$\mathcal{K}_{\mathrm{ti}} = \bigcup_{i=1}^{s} \mathcal{K}_{\mathrm{ti}}^{i} = \bigcup_{i=1}^{s} \bigcup_{j=1}^{r_i} \mathcal{K}_{\mathrm{ti}}^{ij} = \mathcal{R}\left\{\mathbf{V}^{\mathrm{P}}\right\} , \tag{3.28}$$

whereat $\mathbf{V}^{\mathrm{P}}$ denotes the primitive basis of $\mathcal{K}_{\mathrm{ti}}$.

An advantage of using the primitive basis $\mathbf{V}^{\mathrm{P}}$ is the special structure of the corresponding interpolation matrices $\mathbf{S}_V^{\mathrm{P}}$ and $\mathbf{R}^{\mathrm{P}}$, which will be exploited in several proofs later on:

$$\mathbf{S}_V^{\mathrm{P}} = \left[\begin{array}{c|c|c} \begin{matrix}\mathbf{S}_{V,1}^{\mathrm{P}} & & \\ & \ddots & \\ & & \mathbf{S}_{V,i-1}^{\mathrm{P}}\end{matrix} & & \\ \hline & \begin{matrix}\mathbf{S}_{V,i1}^{\mathrm{P}} & & & & & & \\ & \ddots & & & & & \\ & & \mathbf{S}_{V,i(j-1)}^{\mathrm{P}} & & & & \\ & & & \begin{bmatrix} s_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & s_i \end{bmatrix} & & & \\ & & & & \mathbf{S}_{V,i(j+1)}^{\mathrm{P}} & & \\ & & & & & \ddots & \\ & & & & & & \mathbf{S}_{V,ir_i}^{\mathrm{P}}\end{matrix} & \\ \hline & & \begin{matrix}\mathbf{S}_{V,i+1}^{\mathrm{P}} & & \\ & \ddots & \\ & & \mathbf{S}_{V,s}^{\mathrm{P}}\end{matrix} \end{array}\right] , \tag{3.29}$$

$$\mathbf{R}^{\mathrm{P}} = \left[\begin{array}{ccc|ccccccc|ccc} \mathbf{R}_1^{\mathrm{P}} & \cdots & \mathbf{R}_{i-1}^{\mathrm{P}} & \mathbf{R}_{i1}^{\mathrm{P}} & \cdots & \mathbf{R}_{i(j-1)}^{\mathrm{P}} & [\mathbf{r}_{ij} \;\; \mathbf{0} \;\; \cdots \;\; \mathbf{0}] & \mathbf{R}_{i(j+1)}^{\mathrm{P}} & \cdots & \mathbf{R}_{ir_i}^{\mathrm{P}} & \mathbf{R}_{i+1}^{\mathrm{P}} & \cdots & \mathbf{R}_s^{\mathrm{P}} \end{array}\right] .$$

The green and blue highlighted areas in (3.29) are related to s_i (thus $\mathbf{S}_{V,i}^{\mathrm{P}}$ and $\mathbf{R}_i^{\mathrm{P}}$), while the gray entries correspond to all remaining expansion points. The green colored section especially belongs to the j-th tangential direction of s_i (thus $\mathbf{S}_{V,ij}^{\mathrm{P}}$ and $\mathbf{R}_{ij}^{\mathrm{P}}$).

To summarize the results of this section, there are two ways to compute the primitive basis $\mathbf{V}^{\mathrm{P}}$ of an accumulated tangential-input rational Krylov subspace: either by columnwise construction according to Definition 3.6, or by solving the generalized Sylvester equation (3.26). It is left to prove, that (3.26) is solvable with unique solution $\mathbf{V}^{\mathrm{P}}$. For this purpose the solvability of generalized Sylvester equations is discussed in the following section.

3.3 Generalized Sylvester Equations

A Sylvester equation is a matrix equation of the type

$$\mathbf{A}\,\mathbf{X} + \mathbf{X}\,\mathbf{B} = \mathbf{C} \, , \tag{3.30}$$

where the matrices $\mathbf{A} \in \mathbb{C}^{u \times u}$, $\mathbf{B} \in \mathbb{C}^{v \times v}$ and $\mathbf{C} \in \mathbb{C}^{u \times v}$ are assumed to be known and $\mathbf{X} \in \mathbb{C}^{u \times v}$ denotes the (yet unknown) solution [2, p. 173].

Because (3.30) is a linear matrix equation, the solution $\mathbf{X}$ may be obtained by solving a linear system of equations [24]. However, for the purposes of this thesis the numerical solution of Sylvester equations is not needed and therefore not treated.

If the structure of (3.30) is expanded by additional (known) matrices left and right of $\mathbf{X}$, as in

$$\mathbf{A}\,\mathbf{X}\,\mathbf{B} + \mathbf{C}\,\mathbf{X}\,\mathbf{D} = \mathbf{E}\,, \tag{3.31}$$

then the equation is called *generalized* Sylvester equation.

As in the case of a basic linear system of equations ($\mathbf{A}\,\mathbf{x} = \mathbf{b}$), (3.31) may have one, infinitely many or no solutions. In order to show, that a generalized Sylvester equation has an unique solution, following theorem can be used:

Theorem 3.13 (adapted from [12, p. 96])**.** *The matrix equation for* $\mathbf{X} \in \mathbb{F}^{u\times v}$ *(with* $\mathbb{F}$ *as the field* $\mathbb{R}$ *or* $\mathbb{C}$*)*

$$\mathbf{A}\,\mathbf{X}\,\mathbf{B} - \mathbf{C}\,\mathbf{X}\,\mathbf{D} = \mathbf{E}\,, \tag{3.32}$$

where $\mathbf{A}$, $\mathbf{C} \in \mathbb{F}^{u\times u}$ *and* $\mathbf{B}$, $\mathbf{D} \in \mathbb{F}^{v\times v}$*, has a unique solution if and only if*

(i) $(\lambda\,\mathbf{C} - \mathbf{A})$ *and* $(\lambda\,\mathbf{B} - \mathbf{D})$ *are regular matrix pencils, and*

(ii) none of the generalized eigenvalues of the matrix pair $(\mathbf{C},\,\mathbf{A})$ *coincides with the generalized eigenvalues of the matrix pair* $(\mathbf{B},\,\mathbf{D})$*, i. e.* $\lambda\,(\mathbf{C},\,\mathbf{A}) \cap \lambda\,(\mathbf{B},\,\mathbf{D}) = \emptyset$.

Moreover, if all known matrices are real-valued, i. e.

$$\mathbf{A},\,\mathbf{C} \in \mathbb{R}^{u\times u}\,, \qquad \mathbf{B},\,\mathbf{D} \in \mathbb{R}^{v\times v}\,, \qquad \mathbf{E} \in \mathbb{R}^{u\times v} \tag{3.33}$$

then also the solution $\mathbf{X}$ *is real-valued.*

Note that the existence of a unique solution does not depend on the right-hand-side $\mathbf{E}$.

The original theorem in [12] addresses only the case of real-valued matrices. The proof starts with a generalized Schur decomposition, which transforms the matrices left and right of $\mathbf{X}$ to lower and upper triangular form. After that, the special structure of the transformed Sylvester equation is exploited to find the solution of $\mathbf{X}$ and, as a by-product, the conditions for the existence of a unique solution.

Since the generalized Schur decomposition also holds for complex-valued matrices (as claimed in [36, p. 672]) and the rest of the proof gets along with simple matrix multiplications, the original theorem in [12] can be generalized to the complex-valued case in Theorem 3.13.

Example

To illustrate the origin of condition (i) and (ii) of Theorem 3.13, a small example is examined: consider the generalized Sylvester equation

$$\underbrace{\begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{X}} \underbrace{b}_{\mathbf{B}} - \underbrace{\begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{X}} \underbrace{d}_{\mathbf{D}} = \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{\mathbf{E}}\,, \tag{3.34}$$

with the predefined scalars a_1, a_2, b, c, d, e_1, $e_2 \in \mathbb{C}$ and the unknown solution $\mathbf{X} \in \mathbb{C}^{2\times 1}$. According to Definition 2.4 the generalized eigenvalues of the pairs $(\mathbf{C},\,\mathbf{A})$ and $(\mathbf{B},\,\mathbf{D})$ compute to

$$\lambda(\mathbf{C},\,\mathbf{A}) = \{(\mathbb{C},\,0),\,(a_1,\,c)\}\;, \qquad \lambda(\mathbf{B},\,\mathbf{D}) = (d,\,b)\;. \tag{3.35}$$

Separating the rows of (3.34) leads to

$$(a_1\,b - c\,d)\,x_1 = e_1 \qquad \text{and} \qquad (a_2\,b)\,x_2 = e_2\;. \tag{3.36}$$

Obviously a unique solution $\mathbf{X}$ is only possible, if the condition

$$\begin{aligned} b \neq 0 \quad &\Rightarrow \det(\lambda\,\mathbf{B} - \mathbf{D}) = \lambda\,b - d \neq 0 \quad \forall\,\lambda \neq \frac{d}{b} \\ &\Rightarrow (\lambda\,\mathbf{B} - \mathbf{D}) \text{ is regular} \end{aligned} \tag{3.37}$$

holds. Furthermore $a_1\,b \neq c\,d$ is necessary (see (3.36)), which leads together with $b \neq 0$ to[1]

$$\begin{aligned} a_1\,b \neq c\,d \;\text{ and }\; b \neq 0 \quad &\Rightarrow (d,\,b) \neq (a_1,\,c) \;\text{ and }\; (d,\,b) \neq (\mathbb{C},\,0) \\ &\Rightarrow \lambda\,(\mathbf{C},\,\mathbf{A}) \cap \lambda\,(\mathbf{B},\,\mathbf{D}) = \emptyset\;. \end{aligned} \tag{3.38}$$

Additionally the inequality $a_1\,b \neq c\,d$ implies, that either a_1 or c are non-zero. This and the last condition $a_2 \neq 0$ from (3.36) finally show, that

$$\begin{aligned} a_2 \neq 0 \;\text{ and }\; (a_1 \neq 0 \vee c \neq 0) \quad &\Rightarrow \exists\,\lambda \in \mathbb{C} \text{ such that } \det(\lambda\,\mathbf{C} - \mathbf{A}) = (\lambda\,c - a_1)\,a_2 \neq 0 \\ &\Rightarrow (\lambda\,\mathbf{C} - \mathbf{A}) \text{ is regular} \end{aligned} \tag{3.39}$$

holds, thus the conditions of Theorem 3.13 are verified for the given example.

3.4 Equivalence of Rational Krylov Subspaces and Generalized Sylvester Equations

Using the result of the last section one can finally show the equivalence of generating accumulated tangential-input rational Krylov subspaces by construction according to Definition 3.6 and solving the generalized Sylvester equation (3.26). For this purpose Theorem 3.13 is applied to (3.26) using $\mathbf{A} \to \mathbf{A}$, $\mathbf{B} \to \mathbf{I}_q$, $\mathbf{C} \to \mathbf{E}$, $\mathbf{D} \to \mathbf{S}_V^{\mathrm{P}}$ and $\mathbf{X} \to \mathbf{V}^{\mathrm{P}}$, which shows, that in order to get a *unique* solution $\mathbf{V}^{\mathrm{P}}$ of (3.26), three requirements have to be met:

(i) The matrix pencil $\lambda\,\mathbf{E} - \mathbf{A}$ is regular. (This is necessary for the solution of the DAE-system to be unique (see Theorem 2.15), thus this condition is assumed to be fulfilled.)

(ii) The matrix pencil $\lambda\,\mathbf{I}_q - \mathbf{S}_V^{\mathrm{P}}$ is regular. (always fulfilled, see Lemma 2.3)

(iii) None of the generalized eigenvalues of the matrix pair $(\mathbf{E},\,\mathbf{A})$ coincides with the eigenvalues of $\mathbf{S}_V^{\mathrm{P}}$.

To analyze condition (iii) one can make use of the special structure of $\mathbf{S}_V^{\mathrm{P}}$: since $\mathbf{S}_V^{\mathrm{P}}$ is in Jordan canonical form, the eigenvalues are simply the diagonal elements s_i with $i \in \{1,\,2,\,\dots,\,s\}$. That is, condition (iii) demands, that the expansion points s_i differ

[1] Note that generalized eigenvalues are invariant to *equal* scaling of their descriptive scalars, i. e. $(\alpha,\,\beta) = (x\,\alpha,\,x\,\beta)$ for all $x \neq 0$. This directly follows from $\det(\alpha\,\mathbf{X} - \beta\,\mathbf{Y}) = 0 \Leftrightarrow \det(x\,\alpha\,\mathbf{X} - x\,\beta\,\mathbf{Y}) = 0$ with $x \neq 0$.

from the generalized eigenvalues of $(\mathbf{E},\,\mathbf{A})$, which is always assumed for Krylov-based MOR methods. Since in the scope of this thesis expansion points are chosen in the open right half of the complex plane, it is impossible that they coincide with the generalized eigenvalues of an *asymptotically stable* FOM anyway.

The following theorem summarizes the results so far:

Theorem 3.14. *Let*

- *the DAE-system* $(\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C},\,\mathbf{x}_0)$ *describe the FOM where* $\lambda\,\mathbf{E}-\mathbf{A}$ *is regular,*
- $\{s_i\}\subseteq\mathbb{C}$ *be a set of expansion points differing from the generalized eigenvalues of* $(\mathbf{E},\,\mathbf{A})$*,*
- $\{\mathbf{r}_{ij}\}\subseteq\mathbb{C}^{m\times 1}$ *be a set of tangential directions assigned to a specific* s_i*,*
- $q_{ij}\in\mathbb{N}^{>0}$ *be the order of the rational Krylov-subspace related to* s_i *and* $\mathbf{r}_{ij}$*,*
- $\mathcal{K}_{\mathrm{ti}}$ *be the union of all tangential-input rational Krylov subspaces constructed as described in Definition 3.6 and (3.28),*
- $\mathbf{V}^{\mathrm{P}}\in\mathbb{C}^{n\times q}$*,* $\mathbf{S}_V^{\mathrm{P}}\in\mathbb{C}^{q\times q}$ *and* $\mathbf{R}^{\mathrm{P}}\in\mathbb{C}^{m\times q}$ *be the matrices obtained during construction of* $\mathcal{K}_{\mathrm{ti}}$ *as in (3.27), such that the generalized Sylvester equation (3.26) holds for the primitive base* $\mathbf{V}^{\mathrm{P}}$ *of* $\mathcal{K}_{\mathrm{ti}}$*.*

Then the solution $\mathbf{V}^{\mathrm{P}}$ *of (3.26) exists and is* unique.

To check, whether the equivalence also holds for a different base of $\mathcal{K}_{\mathrm{ti}}$, the invertible transformation

$$\mathbf{V}=\mathbf{V}^{\mathrm{P}}\,\mathbf{T}\,,\qquad \mathbf{T}\in\mathbb{C}^{q\times q}\,,\qquad \det(\mathbf{T})\neq 0\,, \tag{3.40}$$

is introduced, with which one can generate any base $\mathbf{V}$ of $\mathcal{K}_{\mathrm{ti}}$. Inserting (3.40) into (3.26) and multiplying with $\mathbf{T}$ from the right leads to

$$\mathbf{A}\,\mathbf{V}-\mathbf{E}\,\mathbf{V}\,\mathbf{S}_V=\mathbf{B}\,\mathbf{R}\,,\quad \text{with } \mathbf{S}_V=\mathbf{T}^{-1}\,\mathbf{S}_V^{\mathrm{P}}\,\mathbf{T} \text{ and } \ \mathbf{R}=\mathbf{R}^{\mathrm{P}}\,\mathbf{T}\,. \tag{3.41}$$

Since the matrices $\mathbf{S}_V^{\mathrm{P}}$ and $\mathbf{S}_V$ are similar, they share the same set of eigenvalues and thus the solution $\mathbf{V}$ remains unique. This leads to the following corollary:

Corollary 3.15. *Let all conditions of Theorem 3.14 hold and let* $\mathbf{V}=\mathbf{V}^{\mathrm{P}}\,\mathbf{T}$ *with* $\mathbf{T}\in\mathbb{C}^{q\times q}$ *denote an arbitrary base of* $\mathcal{K}_{\mathrm{ti}}$*. Then* $\mathbf{V}$ *is the* unique *solution of*

$$\mathbf{A}\,\mathbf{V}-\mathbf{E}\,\mathbf{V}\,\mathbf{S}_V=\mathbf{B}\,\mathbf{R}\,, \tag{3.42}$$

whereby

$$\mathbf{S}_V=\mathbf{T}^{-1}\,\mathbf{S}_V^{\mathrm{P}}\,\mathbf{T}\,,\qquad \mathbf{R}=\mathbf{R}^{\mathrm{P}}\,\mathbf{T}\,. \tag{3.43}$$

Although the eigenvalues of $\mathbf{S}_V^{\mathrm{P}}\rightarrow\mathbf{S}_V$ remain unchanged, it will in general not be in Jordan canonical form anymore. Likewise $\mathbf{R}^{\mathrm{P}}\rightarrow\mathbf{R}$ looses its “beneficial” structure.

Observability of $\mathbf{S}_V$ and $\mathbf{R}$

The assembly of rational Krylov subspaces according to the previous sections delivers a generalized Sylvester equation of special structure: $\mathbf{S}_V^{\mathrm{P}}$ is in Jordan canonical form, while the columns of $\mathbf{R}^{\mathrm{P}}$ are either tangential directions $\mathbf{r}_{ij}$ or zero-vectors. This structure can be exploited to examine the observability of the pair $(\mathbf{S}_V,\, \mathbf{R})$ which is essential for several proofs in the following chapters. For this purpose the well-known criterion by Hautus is used:

Definition 3.16 (adapted from [25, p. 100])**.** The pair $(\mathbf{S}_V^{\mathrm{P}},\, \mathbf{R}^{\mathrm{P}})$ is called (completely) *observable*, if

$$\operatorname{rank}\begin{bmatrix} s_i\,\mathbf{I}_q - \mathbf{S}_V^{\mathrm{P}} \\ \mathbf{R}^{\mathrm{P}} \end{bmatrix} = q\,, \qquad \forall\, s_i \in \lambda(\mathbf{S}_V^{\mathrm{P}}) \tag{3.44}$$

holds.

Exploiting the structure of $\mathbf{S}_V^{\mathrm{P}}$ and $\mathbf{R}^{\mathrm{P}}$ leads to two simple conditions for observability:

Theorem 3.17. *Let $\mathbf{S}_V^{\mathrm{P}}$ and $\mathbf{R}^{\mathrm{P}}$ be as in Theorem 3.14. Then the pair $(\mathbf{S}_V^{\mathrm{P}},\, \mathbf{R}^{\mathrm{P}})$ is observable, if and only if the following conditions hold:*

(i) The tangential directions are non-zero vectors, i. e. $\mathbf{r}_{ij} \neq \mathbf{0}\ \forall\, i,\, j$.

(ii) The tangential directions belonging to an expansion point s_i are linearly independent.

Proof. Since $\mathbf{S}_V^{\mathrm{P}}$ is in Jordan canonical form, it is block-diagonal and the diagonal elements coincide with its eigenvalues. Consider the case of an expansion point s_i with two tangential directions $\mathbf{r}_{i1}$, $\mathbf{r}_{i2}$ of order $q_{i1} = 2$ and $q_{i2} = 1$ respectively:

$$\operatorname{rank}\begin{bmatrix} s_i\,\mathbf{I}_q - \mathbf{S}_V^{\mathrm{P}} \\ \mathbf{R}^{\mathrm{P}} \end{bmatrix} =$$

$$= \operatorname{rank}\left[\begin{array}{ccc|cc|ccc}
\left(s_i\,\mathbf{I}_{q_1} - \mathbf{S}_{V,1}^{\mathrm{P}}\right) & & & & & & & \\
 & \ddots & & & & & & \\
 & & \left(s_i\,\mathbf{I}_{q(i-1)} - \mathbf{S}_{V,i-1}^{\mathrm{P}}\right) & & & & & \\ \hline
 & & & \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} & & & & \\
 & & & & 0 & & & \\ \hline
 & & & & & \left(s_i\,\mathbf{I}_{q(i+1)} - \mathbf{S}_{V,i+1}^{\mathrm{P}}\right) & & \\
 & & & & & & \ddots & \\
 & & & & & & & \left(s_i\,\mathbf{I}_{q_s} - \mathbf{S}_{V,s}^{\mathrm{P}}\right) \\ \hline
\cdots & \cdots & \cdots & \begin{bmatrix} \mathbf{r}_{i1} & \mathbf{0} \end{bmatrix} & \mathbf{r}_{i2} & \cdots & \cdots & \cdots
\end{array}\right] \tag{3.45}$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{(q+m)\times\left(\sum\limits_{u=1}^{i-1} q_u\right)} \quad \underbrace{\qquad\quad}_{(q+m)\times 3} \quad \underbrace{\qquad\qquad\qquad\qquad}_{(q+m)\times\left(\sum\limits_{u=i+1}^{s} q_u\right)}$$

White areas denote zero entries, while colored blocks may contain non-zero values. All green highlighted matrices belong to the first tangential direction $\mathbf{r}_{i1}$ and occupy $q_{i1} = 2$ columns. The same holds for the blue colored blocks related to $\mathbf{r}_{i2}$.

Because $s_i \neq s_w \ \forall \ i \neq w$ all eigenvalues of the Jordan-blocks $\mathbf{S}^{\mathrm{P}}_{V,1}, \dots, \mathbf{S}^{\mathrm{P}}_{V,i-1}$ and $\mathbf{S}^{\mathrm{P}}_{V,i+1}, \dots, \mathbf{S}^{\mathrm{P}}_{V,s}$ differ from s_i. Therefore the upper gray part of the overall matrix has non-zero elements on its diagonal, thus all columns containing gray areas are linearly independent (to the green/blue part and to each other).

It is left to show, that the columns containing green and blue areas are linearly independent, which is obviously only the case, if condition (i) and (ii) are fulfilled. ■

The connection of $(\mathbf{S}^{\mathrm{P}}_V, \mathbf{R}^{\mathrm{P}})$ and $(\mathbf{S}_V, \mathbf{R})$ through the *invertible* matrix $\mathbf{T}$ finally allows following statement:

Corollary 3.18. *Let $\mathbf{T} \in \mathbb{C}^{q\times q}$ be a regular transformation matrix and let $\mathbf{S}_V$ and $\mathbf{R}$ be as defined in (3.43). Then the pair $(\mathbf{S}_V, \mathbf{R})$ is observable, if and only if the pair $(\mathbf{S}^{\mathrm{P}}_V, \mathbf{R}^{\mathrm{P}})$ is observable.*

Proof. Consider the transformation matrices

$$\mathbf{T}_l := \begin{bmatrix} \mathbf{T}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \in \mathbb{C}^{(q+m)\times(q+m)} \qquad \text{and} \qquad \mathbf{T}_r := \mathbf{T} \in \mathbb{C}^{q\times q} \,. \tag{3.46}$$

Since both, $\mathbf{T}_l$ and $\mathbf{T}_r$, are invertible, the equality

$$\operatorname{rank}\begin{bmatrix} s_i\,\mathbf{I}_q - \mathbf{S}^{\mathrm{P}}_V \\ \mathbf{R}^{\mathrm{P}} \end{bmatrix} = \operatorname{rank}\left(\mathbf{T}_l \begin{bmatrix} s_i\,\mathbf{I}_q - \mathbf{S}^{\mathrm{P}}_V \\ \mathbf{R}^{\mathrm{P}} \end{bmatrix} \mathbf{T}_r\right) = \operatorname{rank}\begin{bmatrix} s_i\,\mathbf{I}_q - \mathbf{S}_V \\ \mathbf{R} \end{bmatrix} \tag{3.47}$$

holds (see [32, p. 9]). ■

Note that both conditions of Theorem 3.17 are common design rules in MOR, since a violation would result in a rank deficient $\mathbf{V}$. Therefore it is assumed in the following, that the conditions of Theorem 3.17 are fulfilled such that $(\mathbf{S}_V, \mathbf{R})$ is observable.

3.5 Model Order Reduction of DAEs

In the following the concept of *projective* MOR is introduced and conditions for tangential interpolation are presented. Although the $\mathcal{H}_2$ pseudo-optimal reduction scheme discussed in Chapter 4 does not directly belong to projective MOR, there are strong relations (and even equivalence, if certain conditions are met). The proof of tangential interpolation in the case of PORK is postponed to Chapter 4.

3.5.1 Projective MOR and Tangential Interpolation with Rational Krylov Subspace Methods

As mentioned in Chapter 1, the most popular methods for the reduction of LTI-systems are BT and rational Krylov subspace methods. Both belong to the field of *projective* MOR, as they project the FOM onto a defined subspace in order to obtain the ROM. Depending on the applied technique, different projections are performed.

The main goal of MOR is to change the number of degrees of freedom from n (order of the FOM) to q (order of the ROM) with $q \ll n$. For this purpose the system state of the FOM $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ is approximated through $\mathbf{x} \approx \mathbf{V}\mathbf{x}_\mathrm{r}$, where $\mathbf{V} \in \mathbb{R}^{n\times q}$ denotes the "first" projection matrix:

$$\mathbf{E}\,\mathbf{V}\,\dot{\mathbf{x}}_\mathrm{r} = \mathbf{A}\,\mathbf{V}\,\mathbf{x}_\mathrm{r} + \mathbf{B}\,\mathbf{u} + \underbrace{(\mathbf{A}\,\mathbf{x} - \mathbf{A}\,\mathbf{V}\,\mathbf{x}_\mathrm{r} - \mathbf{E}\,\dot{\mathbf{x}} + \mathbf{E}\,\mathbf{V}\,\dot{\mathbf{x}}_\mathrm{r})}_{\boldsymbol{\epsilon}} \,. \tag{3.48}$$

Herein $\boldsymbol{\epsilon}$ is a residuum which contains the error of the system dynamics caused by the approximation. If a "second" projection matrix $\mathbf{W} \in \mathbb{R}^{n\times q}$ is multiplied from the left and $\mathbf{W}^\mathrm{T}\,\boldsymbol{\epsilon} = \mathbf{0}$ is enforced (*Petrov-Galerkin*-condition, [2, p. 279]) one obtains

$$\mathbf{W}^\mathrm{T}\,\mathbf{E}\,\mathbf{V}\,\dot{\mathbf{x}}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{A}\,\mathbf{V}\,\mathbf{x}_\mathrm{r} + \mathbf{W}^\mathrm{T}\,\mathbf{B}\,\mathbf{u} \,. \tag{3.49}$$

The reduced output finally reads as $\mathbf{y}_\mathrm{r} = \mathbf{C}\,\mathbf{V}\,\mathbf{x}_\mathrm{r}$ such that a realization of the ROM is given by $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ with

$$\mathbf{E}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{E}\,\mathbf{V}\,, \quad \mathbf{A}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{A}\,\mathbf{V}\,, \quad \mathbf{B}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{B}\,, \quad \mathbf{C}_\mathrm{r} = \mathbf{C}\,\mathbf{V}\,. \tag{3.50}$$

A graphical interpretation of the projection is depicted in Figure 3.4. Therein $n = 3$ (thus $\mathbf{x}(t) \in \mathbb{R}^3$) and $q = 2$ with $\mathbf{V} = [\mathbf{e}_1,\, \mathbf{e}_2]$ in combination with orthogonal projection (i. e. $\mathbf{W} = \mathbf{V}$) is chosen for simplicity.

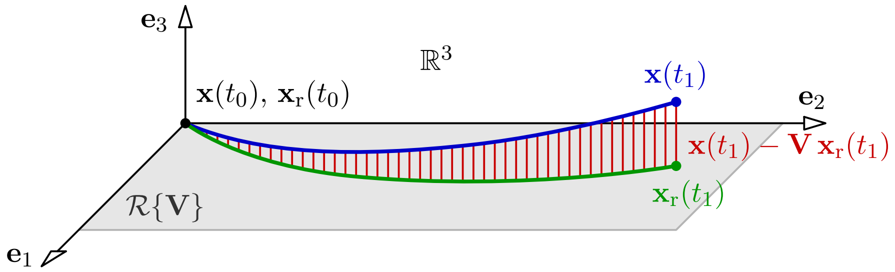

$\mathbf{e}_3$
$\mathbb{R}^3$
$\mathbf{x}(t_0),\, \mathbf{x}_\mathrm{r}(t_0)$
$\mathbf{x}(t_1)$
$\mathbf{e}_2$
$\mathbf{x}(t_1) - \mathbf{V}\,\mathbf{x}_\mathrm{r}(t_1)$
$\mathbf{x}_\mathrm{r}(t_1)$
$\mathcal{R}\{\mathbf{V}\}$
$\mathbf{e}_1$

Figure 3.4: Illustration of projective MOR for the case of orthogonal projection with $n = 3$, $q = 2$ and initialization at $\mathbf{x}(t_0) = \mathbf{x}_\mathrm{r}(t_0) = \mathbf{0}$. The trajectory of the FOM (blue) is projected onto $\mathbf{V}$ (gray plane). The resulting approximation $\mathbf{x}_\mathrm{r}(t)$ is highlighted in green, while the error $\mathbf{x}(t) - \mathbf{V}\,\mathbf{x}_\mathrm{r}(t)$ is colored red.

The main difficulty in projective MOR is to find appropriate choices for $\mathbf{V}$ and $\mathbf{W}$ in order to obtain good approximation results. In the case of rational Krylov subspace methods, one (*one-sided reduction*) or both (*two-sided reduction*) projection matrices span accumulated rational Krylov subspaces. This is because the special choice $\mathcal{K}_\mathrm{ti} = \mathcal{R}\,\{\mathbf{V}\}$ and/or $\mathcal{K}_\mathrm{to} = \mathcal{R}\,\{\mathbf{W}\}$ leads to tangential interpolation:

Key Theorem 3.19 (adapted from [3, p. 10])**.** *Let*

- $\mathbf{G}(s)$ *be the transfer function of the FOM which is described by the DAE-system* $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ *and*
- $\mathbf{G}_\mathrm{r}(s)$ *be the transfer function of the ROM which is obtained by projection with* $\mathbf{V}$ *and* $\mathbf{W}$ *according to (3.50).*

(i) If

$$\Big[(\mathbf{A} - s_i\,\mathbf{E})^{-1}\,\mathbf{E}\Big]^\mu (\mathbf{A} - s_i\,\mathbf{E})^{-1}\,\mathbf{B}\,\mathbf{r}_{ij} \in \mathcal{R}(\mathbf{V}) \quad \forall\, \mu = 0,\, \dots,\, q_{ij} - 1\,, \tag{3.51}$$

or equivalently, if

$$\mathcal{K}_{\mathrm{ti}}^{ij}\left(s_i,\, \mathbf{r}_{ij},\, q_{ij}\right) \subseteq \mathcal{R}(\mathbf{V})\ , \tag{3.52}$$

where $s_i \in \mathbb{C}$ *such that* $(\mathbf{A} - s_i\, \mathbf{E})$ *and* $(\mathbf{A}_\mathrm{r} - s_i\, \mathbf{E}_\mathrm{r})$ *are invertible,* $\mathbf{r}_{ij} \neq \mathbf{0}$ *and* $q_{ij} \in \mathbb{N}^{>0}$*, then*

$$\left(\frac{\mathrm{d}^\mu \mathbf{G}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} = \left(\frac{\mathrm{d}^\mu \mathbf{G}_\mathrm{r}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} \quad \forall\, \mu = 0,\, \dots,\, q_{ij} - 1\ . \tag{3.53}$$

(ii) If

$$\left[(\mathbf{A} - s_i\, \mathbf{E})^{-\mathrm{T}}\ \mathbf{E}^\mathrm{T}\right]^\mu (\mathbf{A} - s_i\, \mathbf{E})^{-\mathrm{T}}\ \mathbf{C}^\mathrm{T}\, \mathbf{l}_{ij} \in \mathcal{R}(\mathbf{W}) \quad \forall\, \mu = 0,\, \dots,\, q_{ij} - 1\ , \tag{3.54}$$

or equivalently, if

$$\mathcal{K}_{\mathrm{to}}^{ij}\left(s_i,\, \mathbf{l}_{ij},\, q_{ij}\right) \subseteq \mathcal{R}(\mathbf{W})\ , \tag{3.55}$$

where $s_i \in \mathbb{C}$ *such that* $(\mathbf{A} - s_i\, \mathbf{E})$ *and* $(\mathbf{A}_\mathrm{r} - s_i\, \mathbf{E}_\mathrm{r})$ *are invertible,* $\mathbf{l}_{ij} \neq \mathbf{0}$ *and* $q_{ij} \in \mathbb{N}^{>0}$*, then*

$$\mathbf{l}_{ij}^\mathrm{T} \cdot \left(\frac{\mathrm{d}^\mu \mathbf{G}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} = \mathbf{l}_{ij}^\mathrm{T} \cdot \left(\frac{\mathrm{d}^\mu \mathbf{G}_\mathrm{r}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \quad \forall\, \mu = 0,\, \dots,\, q_{ij} - 1\ . \tag{3.56}$$

(iii) If both of the previous statements hold for the same s_i*, then*

$$\mathbf{l}_{ij}^\mathrm{T} \cdot \left(\frac{\mathrm{d}^\mu \mathbf{G}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} = \mathbf{l}_{ij}^\mathrm{T} \cdot \left(\frac{\mathrm{d}^\mu \mathbf{G}_\mathrm{r}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} \quad \forall\, \mu = 0,\, \dots,\, 2\, q_{ij} - 1\ . \tag{3.57}$$

Remark 3.20. Note that there are two small typos in (b) and (c) of [3, theorem 2]: first, the vector b is misplaced in (b) and second, the range of l should be $0, \dots, M + N - 1$ in (c).

According to Definition 3.2 the relations (3.53) and (3.56) are descriptions of *right* and *left* tangential interpolation respectively. Furthermore (3.57) corresponds to *two-sided* reduction which causes twice as much moments to be matched.

According to Theorem 3.19 the results of Sections 3.2 and 3.4 can be used to describe the process of tangential interpolation in the form of generalized Sylvester equations, i. e. the interpolation data (expansion points s_i, tangential directions $\mathbf{r}_{ij}$ and order q_{ij}) are encoded in the matrices $\mathbf{S}_V$ and $\mathbf{R}$ of (3.42). Note that Theorem 3.19 makes use of the images of $\mathbf{V}$ and $\mathbf{W}$, thus the actual bases are irrelevant.

Remark 3.21. In the previous sections only tangential-input rational Krylov subspaces are handled. Due to duality all results can be transferred to the case of tangential-output rational Krylov subspaces. This leads to the generalized Sylvester equation

$$\mathbf{A}^\mathrm{T}\, \mathbf{W} - \mathbf{E}^\mathrm{T}\, \mathbf{W}\, \mathbf{S}_W = \mathbf{C}^\mathrm{T}\, \mathbf{L}\ , \tag{3.58}$$

wherein $\mathbf{L} \in \mathbb{C}^{p \times q}$ contains the set of *left* tangential directions $\{\mathbf{l}_{ij}\}$ analogous to the assembly of $\{\mathbf{r}_{ij}\}$ to $\mathbf{R}$ and $\mathbf{S}_W$ serves the same purpose as $\mathbf{S}_V$.

3.5.2 General Framework for the Reduction of Improper DAEs

As exemplified in [18, example 2.1], basic tangential interpolation is not sufficient during the reduction of (improper) DAEs. This is because $\mathbf{W}^{\mathrm{T}}\,\mathbf{E}\,\mathbf{V} = \mathbf{E}_{\mathrm{r}}$ will generically be regular (⇒ ROM is of ODE-type), even if the FOM is an improper DAE [18, p. B1013]. Although moment matching at predefined expansion points is guaranteed, the frequency responses diverge since the polynomial part of the FOM dominates at high frequencies (see Figure 3.1). This causes the overall error to be unbounded, which is not tolerable in most applications. Therefore it is important to keep the improper part of the transfer function unchanged, i. e. the reduction process may only affect the slow subsystem.

One of the goals of this thesis is to port the PORK algorithm to the DAE-case. Since $\mathcal{H}_2$ pseudo-optimality as the result of PORK bases on the $\mathcal{H}_2$ inner-product, which is only defined for strictly proper transfer functions (see Section 4.1), a partitioning of the FOM is necessary. This is done by computing realizations of the strictly proper part $\mathbf{G}^{\mathrm{sp}}(s)$ and the improper part $\mathbf{P}(s)$ of the transfer function[2]. At this point the structure of the DAE is exploited: since it is assumed that the spectral projectors are known in advance, it is possible to calculate strictly proper and improper realizations in an inexpensive way (see Corollary 2.23). The partitioning allows to process the strictly proper subsystem separately which is presented in Chapter 4.

Although $\mathbf{P}(s)$ has to be fit exactly, some kind of "reduction" is necessary for the improper subsystem too: computing a realization of $\mathbf{P}(s)$ according to Corollary 2.23 projects the matrix $\mathbf{B}$ and/or $\mathbf{C}$. In contrast the full-dimensional matrices $\mathbf{E}$ and $\mathbf{A}$ stay the same. Without any modification a concatenation with the reduced strictly proper subsystem would result in dim(ROM) > dim(FOM), which is exactly the opposite of what MOR is meant for. Thus the improper subsystem has to be reformulated, while keeping its contribution to the transfer function exactly the same. This can be achieved in an elegant way by exploiting the structure of the DAE again to find a minimal realization of $\mathbf{P}(s)$, which is discussed in Chapter 5.

After all the reduced and reformulated subsystems are combined to the final ROM (which is of DAE-type). This is done by simple addition of the transfer functions, i. e. $\mathbf{G}_{\mathrm{r}}(s) = \mathbf{G}_{\mathrm{r}}^{\mathrm{sp}}(s) + \mathbf{P}_{\mathrm{r}}(s)$, which is equivalent to the assembly

$$\underbrace{\begin{bmatrix} \mathbf{E}_{\mathrm{r}}^{\mathrm{sp}} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{\mathrm{r}}^{\mathrm{im}} \end{bmatrix}}_{\mathbf{E}_{\mathrm{r}}} \underbrace{\begin{bmatrix} \dot{\mathbf{x}}_{\mathrm{r}}^{\mathrm{sp}} \\ \dot{\mathbf{x}}_{\mathrm{r}}^{\mathrm{im}} \end{bmatrix}}_{\dot{\mathbf{x}}_{\mathrm{r}}} = \underbrace{\begin{bmatrix} \mathbf{A}_{\mathrm{r}}^{\mathrm{sp}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{\mathrm{r}}^{\mathrm{im}} \end{bmatrix}}_{\mathbf{A}_{\mathrm{r}}} \underbrace{\begin{bmatrix} \mathbf{x}_{\mathrm{r}}^{\mathrm{sp}} \\ \mathbf{x}_{\mathrm{r}}^{\mathrm{im}} \end{bmatrix}}_{\mathbf{x}_{\mathrm{r}}} + \underbrace{\begin{bmatrix} \mathbf{B}_{\mathrm{r}}^{\mathrm{sp}} \\ \mathbf{B}_{\mathrm{r}}^{\mathrm{im}} \end{bmatrix}}_{\mathbf{B}_{\mathrm{r}}} \mathbf{u}\,, \quad \mathbf{y}_{\mathrm{r}} = \underbrace{\begin{bmatrix} \mathbf{C}_{\mathrm{r}}^{\mathrm{sp}} & \mathbf{C}_{\mathrm{r}}^{\mathrm{im}} \end{bmatrix}}_{\mathbf{C}_{\mathrm{r}}} \underbrace{\begin{bmatrix} \mathbf{x}_{\mathrm{r}}^{\mathrm{sp}} \\ \mathbf{x}_{\mathrm{r}}^{\mathrm{im}} \end{bmatrix}}_{\mathbf{x}_{\mathrm{r}}}\,. \tag{3.59}$$

where $[\mathbf{E}_{\mathrm{r}}^{\mathrm{sp}}, \mathbf{A}_{\mathrm{r}}^{\mathrm{sp}}, \mathbf{B}_{\mathrm{r}}^{\mathrm{sp}}, \mathbf{C}_{\mathrm{r}}^{\mathrm{sp}}]$ and $[\mathbf{E}_{\mathrm{r}}^{\mathrm{im}}, \mathbf{A}_{\mathrm{r}}^{\mathrm{im}}, \mathbf{B}_{\mathrm{r}}^{\mathrm{im}}, \mathbf{C}_{\mathrm{r}}^{\mathrm{im}}]$ denote realizations of the reduced strictly proper and reformulated improper subsystems respectively.

An illustration of the proposed overall framework for the reduction of improper DAEs is given in Figure 3.5. Note that in several technical applications the FOM is strictly proper (even though formulated as a DAE). If this circumstance is known in advance, no partitioning into subsystems is necessary. Instead it is sufficient to follow the left path in Figure 3.5. Nevertheless a projection onto the deflating subspace corresponding to the finite eigenvalues may still be necessary, in order to avoid numerical issues.

[2] Keep in mind that $\mathbf{G}(s) = \mathbf{G}^{\mathrm{sp}}(s) + \mathbf{P}(s)$ holds.

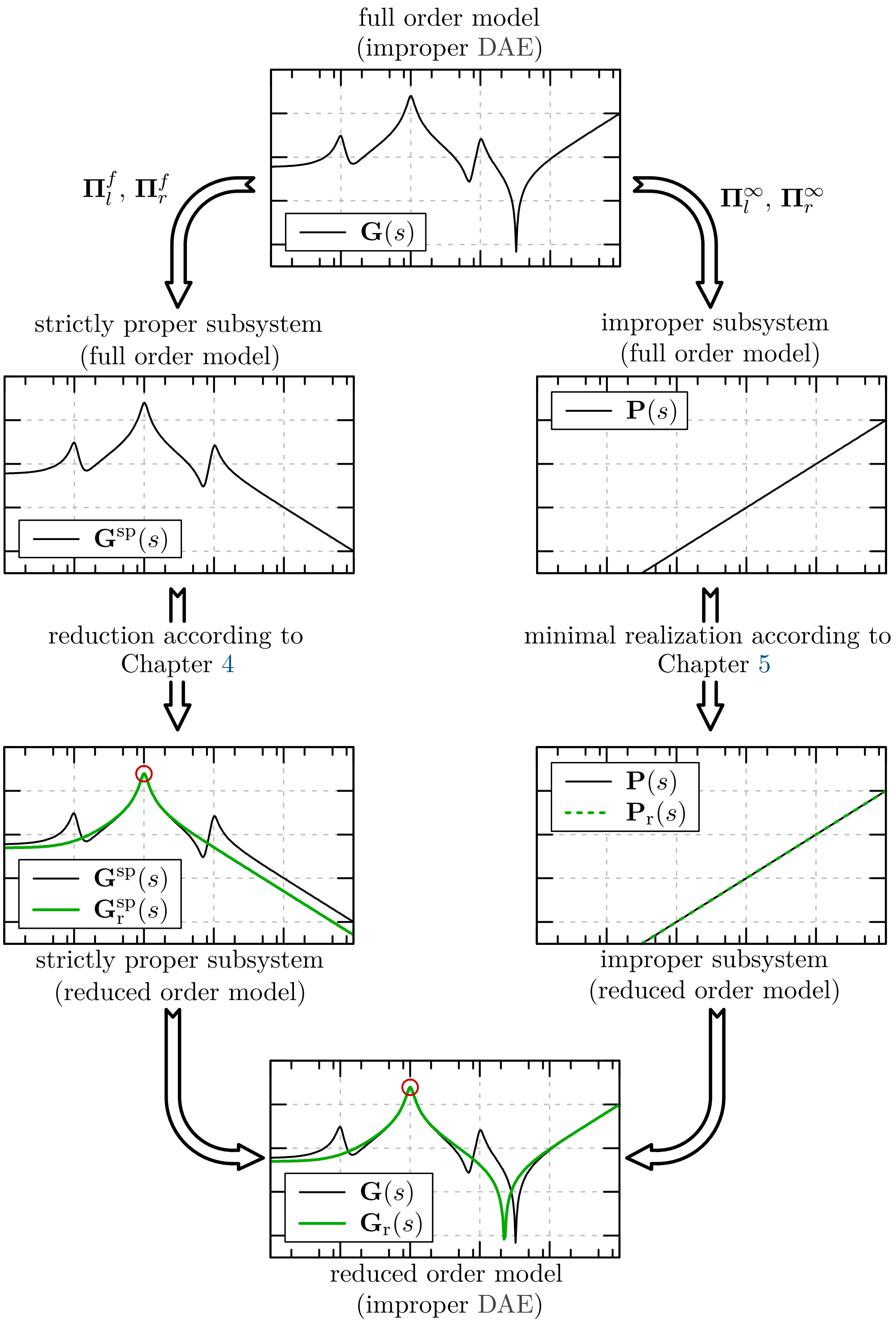

full order model
(improper DAE)
$\mathbf{G}(s)$
$\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_r^f$
$\mathbf{\Pi}_l^\infty$, $\mathbf{\Pi}_r^\infty$
strictly proper subsystem
(full order model)
$\mathbf{G}^{\mathrm{sp}}(s)$
improper subsystem
(full order model)
$\mathbf{P}(s)$
reduction according to
Chapter 4
minimal realization according to
Chapter 5
$\mathbf{G}^{\mathrm{sp}}(s)$
$\mathbf{G}_{\mathrm{r}}^{\mathrm{sp}}(s)$
$\mathbf{P}(s)$
$\mathbf{P}_{\mathrm{r}}(s)$
strictly proper subsystem
(reduced order model)
improper subsystem
(reduced order model)
$\mathbf{G}(s)$
$\mathbf{G}_{\mathrm{r}}(s)$
reduced order model
(improper DAE)

Figure 3.5: General framework for the reduction of improper DAEs: the left side deals with the reduction of the strictly proper subsystem, while the right part depicts the transformation of the improper subsystem into its minimal realization. A red circle indicates moment matching. Note that since $\mathbf{P}(s)$ is fit exactly (i. e. $\mathbf{P}(s) = \mathbf{P}_{\mathrm{r}}(s)$), tangential interpolation of the strictly proper subsystem ($\mathbf{G}^{\mathrm{sp}}(s) \leftrightarrow \mathbf{G}_{\mathrm{r}}^{\mathrm{sp}}(s)$) also applies to the overall ROM ($\mathbf{G}(s) \leftrightarrow \mathbf{G}_{\mathrm{r}}(s)$).

Chapter 4

Adaptive Reduction of the Strictly Proper Subsystem

In the following the reduction process of the strictly proper subsystem is presented. For this purpose it is assumed, that either the DAE describing the FOM is strictly proper by itself, or a projection of **B** or **C** according to Corollary 2.23 has been performed as a preprocessing step. For the ease of notation the superscript "sp" (indicating the relation to the strictly proper subsystem) is omitted within this chapter.

First a formulation of the $\mathcal{H}_2$ inner-product of DAEs is derived, which is a general result and independent of MOR. Within the scope of this thesis it serves as an important tool to prove, that the PORK algorithm, originally introduced for the ODE-case, is applicable *without any modifications* for strictly proper DAE-systems. Finally adaptive MOR-techniques associated with PORK, namely the SPARK algorithm and the CURE framework, are discussed in Section 4.3. The term "adaptive" refers therein to an automatic selection of the interpolation data (with SPARK) and the order of the ROM (with CURE).

As stated above, the actual reduction is done using the PORK algorithm, which leads to an $\mathcal{H}_2$ pseudo-optimal ROM. However the concept of $\mathcal{H}_2$ pseudo-optimality has not been explained so far. Since important fundamentals, which are necessary to understand $\mathcal{H}_2$ pseudo-optimality, have not been presented yet, the precise definition is postponed to Section 4.2. As a short anticipation, one might summarize $\mathcal{H}_2$ pseudo-optimality as follows: instead of analyzing the set of all possible ROMs to find the optimum (with respect to the $\mathcal{H}_2$ norm of the error system), one restricts the search to a predefined subset. Since the obtained ROM is only optimal within this specific subspace, the term "pseudo" is used. Note that the SPARK algorithm discussed in Section 4.3 automatically selects an optimal subspace, such that (local) $\mathcal{H}_2$ optimal and $\mathcal{H}_2$ pseudo-optimal ROM coincide.

4.1 $\mathcal{H}_2$ Inner-Product of Strictly Proper DAEs

This section presents a formulation of the $\mathcal{H}_2$ inner-product of two strictly proper DAE-systems (or more precisely: of their transfer functions) using generalized Sylvester equations. Since the following results are not directly related to MOR, the terms FOM and ROM will not be used. Instead two *asymptotically stable* and *strictly proper* DAE-systems with the transfer functions $\mathbf{G}(s) \in \mathbb{C}^{p\times m}$ and $\mathbf{G}_{\mathrm{H}}(s) \in \mathbb{C}^{p\times m}$ are considered. The corresponding realizations used for analysis are denoted by $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ and $[\mathbf{E}_{\mathrm{H}},\, \mathbf{A}_{\mathrm{H}},\, \mathbf{B}_{\mathrm{H}},\, \mathbf{C}_{\mathrm{H}}]$ respectively.

4.1.1 Fundamentals

First of all the $\mathcal{H}_2$ space as a subspace of $\mathcal{L}_2$ is defined:

Definition 4.1 (adapted from [40, p. 845] and [7, p. 4953])**.** Let $\mathcal{L}_2^{(p,m)}$ be the Hilbert space of matrix-valued functions $\mathbf{F} : \imath\,\mathbb{R} \to \mathbb{C}^{p\times m}$ that have bounded $\mathcal{L}_2^{(p,m)}$-norm

$$\|\mathbf{F}\|_{\mathcal{L}_2^{(p,m)}} := \left(\frac{1}{2\,\pi}\int_{-\infty}^{\infty} \|\mathbf{F}(\imath\,\omega)\|_F^2 \,\mathrm{d}\omega\right)^{\frac{1}{2}}, \tag{4.1}$$

wherein $\|\cdot\|_F$ denotes the Frobenius norm. Then $\mathcal{H}_2^{(p,m)}$ denotes the subspace of $\mathcal{L}_2^{(p,m)}$ containing all rational functions with $\mathcal{O}(\text{numerator}) < \mathcal{O}(\text{denominator})$ that are analytic in the closed right half of the complex plane, i. e. they have only poles in the open left half of the complex plane. A function which is contained in $\mathcal{H}_2^{(p,m)}$ is called $\mathcal{H}_2^{(p,m)}$-function.

Note that m and p in the superscript "(p,m)" are related to the count of inputs and outputs of the system. The definition of the $\mathcal{H}_2$-function space directly allows to determine the behavior at very high frequencies:

Corollary 4.2. *Let* $\mathbf{F} : \imath\,\mathbb{R} \to \mathbb{C}^{p\times m}$ *be a* $\mathcal{H}_2^{(p,m)}$*-function as specified in Definition 4.1. Then* $\lim\limits_{\omega\to\infty} \mathbf{F}(\imath\,\omega) = \mathbf{0}$ *holds.*

Proof. According to Definition 4.1 $\mathbf{F}(\imath\,\omega)$ is a rational function with $\mathcal{O}(\text{numerator}) < \mathcal{O}(\text{denominator})$. Therefore the denominator grows faster than the numerator for increasing ω which proves the claim. ■

As mentioned above, there are specific requirements on transfer functions in order to comply with the $\mathcal{H}_2$ inner-product. This is justified in the following lemma:

Lemma 4.3 (adapted from [40, p. 845])**.** *Let* $\mathbf{G}(s) \in \mathbb{C}^{p\times m}$ *be the transfer function of a DAE-system. Then following statements hold:*

(i) If the system is asymptotically stable and strictly proper, then $\mathbf{G}(s)$ *is a* $\mathcal{H}_2^{(p,m)}$*-function, i. e.* $\mathbf{G}(s) \in \mathcal{H}_2^{(p,m)}$*.*

(ii) If the system is improper (or proper), then $\mathbf{G}(s)$ *is not a* $\mathcal{H}_2^{(p,m)}$*-function, i. e.* $\mathbf{G}(s) \notin \mathcal{H}_2^{(p,m)}$*.*

Note that the statements of Lemma 4.3 are independent of each other, because asymptotic stability (in combination with strictly properness) is a sufficient but *not necessary* condition, i. e. $\mathbf{G}(s) \in \mathcal{H}_2^{(p,m)}$ does not imply $\mathrm{Re}\{\lambda_f(\mathbf{E},\, \mathbf{A})\} < 0$ [40, p. 845]. This is related to the difference between the (finite) eigenvalues of $(\mathbf{E},\, \mathbf{A})$ and the poles of $\mathbf{G}(s)$.

Making use of Corollary 2.31, a statement about minimality of $\mathcal{H}_2^{(p,m)}$-functions in the context of DAEs is possible:

Lemma 4.4. *Let* $\mathbf{G}(s) \in \mathbb{C}^{p\times m}$ *be the transfer function of a DAE-system. If* $\mathbf{G}(s)$ *is a* $\mathcal{H}_2^{(p,m)}$*-function, then it is not minimal and admits a realization* $[\hat{\mathbf{E}}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}]$ *in ODE-form, i. e.* $\det(\hat{\mathbf{E}}) \neq 0$.

Proof. Since $\mathbf{G}(s) \in \mathcal{H}_2^{(p,m)}$ holds, it is strictly proper according to Lemma 4.3 and thus Corollary 2.31 on the existence of such ODE-realization can be applied. ■

Finally the $\mathcal{H}_2$ inner-product is formulated in its most general form:

Definition 4.5 ([7, p. 4953])**.** The $\mathcal{H}_2$ inner-product of two $\mathcal{H}_2^{(p,m)}$-functions $\mathbf{G}(s) \in \mathbb{C}^{p\times m}$ and $\mathbf{G}_{\mathrm{H}}(s) \in \mathbb{C}^{p\times m}$ is defined as

$$\begin{aligned}\langle \mathbf{G}, \mathbf{G}_{\mathrm{H}}\rangle_{\mathcal{H}_2^{(p,m)}} &:= \frac{1}{2\,\pi}\int_{-\infty}^{\infty} \mathrm{tr}\left\{\mathbf{G}_{\mathrm{H}}(\imath\,\omega)\,\mathbf{G}^*(\imath\,\omega)\right\}\mathrm{d}\omega \\ &= \text{... Parseval's relation [45, p. 148] ...} = \int_0^{\infty} \mathrm{tr}\left\{\mathbf{G}_{\mathrm{H}}(t)\,\mathbf{G}^*(t)\right\}\mathrm{d}t\,.\end{aligned} \tag{4.2}$$

Note that $\mathbf{G}(s)$ and $\mathbf{G}_{\mathrm{H}}(s)$ must have the same dimensions $p \times m$, i. e. the same count of inputs and outputs, in order to be combined in $\langle \mathbf{G}, \mathbf{G}_{\mathrm{H}}\rangle_{\mathcal{H}_2^{(p,m)}}$. Furthermore (4.2) shows, that the $\mathcal{H}_2$ inner-product may be formulated either in the frequency or time domain. Therefore the parameters "(s)" and "(t)" are omitted in $\langle \mathbf{G}, \mathbf{G}_{\mathrm{H}}\rangle_{\mathcal{H}_2^{(p,m)}}$. Using the definition of the inner-product, the $\mathcal{H}_2$ norm is derived:

Definition 4.6 ([7, p. 4953])**.** The $\mathcal{H}_2$ norm of a $\mathcal{H}_2^{(p,m)}$-function $\mathbf{G}(s) \in \mathbb{C}^{p\times m}$ is defined as

$$\|\mathbf{G}\|^2_{\mathcal{H}_2^{(p,m)}} := \frac{1}{2\,\pi}\int_{-\infty}^{\infty} \|\mathbf{G}(\imath\,\omega)\|_F^2 \mathrm{d}\omega = \langle \mathbf{G}, \mathbf{G}\rangle_{\mathcal{H}_2^{(p,m)}}\,. \tag{4.3}$$

In its most general form, the $\mathcal{H}_2$ inner-product is not commutative, i. e. the order of its members is important. Fortunately a restriction to technical applications (or more precisely to real-valued systems) ensures commutativity of (4.2):

Lemma 4.7. *Let* $\mathbf{G}(s)$ *and* $\mathbf{G}_{\mathrm{H}}(s)$ *be the transfer functions of two asymptotically stable and strictly proper DAE-systems which allow realizations with* real-valued *system matrices. Then the equality*

$$\langle \mathbf{G}, \mathbf{G}_{\mathrm{H}}\rangle_{\mathcal{H}_2^{(p,m)}} = \langle \mathbf{G}_{\mathrm{H}}, \mathbf{G}\rangle_{\mathcal{H}_2^{(p,m)}} \in \mathbb{R} \tag{4.4}$$

holds.

Proof. According to Remark 2.27 the impulse responses $\mathbf{G}(t)$ and $\mathbf{G}_{\mathrm{H}}(t)$ have to be real-valued. This leads together with (4.2) to

$$\begin{aligned}\langle \mathbf{G}, \mathbf{G}_{\mathrm{H}}\rangle_{\mathcal{H}_2^{(p,m)}} &= \int_0^{\infty} \mathrm{tr}\left\{\mathbf{G}_{\mathrm{H}}(t)\,\mathbf{G}^{\mathrm{T}}(t)\right\}\mathrm{d}t = \int_0^{\infty} \mathrm{tr}\left\{\mathbf{G}(t)\,\mathbf{G}_{\mathrm{H}}^{\mathrm{T}}(t)\right\}\mathrm{d}t \\ &= \langle \mathbf{G}_{\mathrm{H}}, \mathbf{G}\rangle_{\mathcal{H}_2^{(p,m)}}\,.\end{aligned} \tag{4.5}$$

■

4.1.2 $\mathcal{H}_2$ Inner-Product as the Solution of a Generalized Sylvester Equation

The following investigations are strongly related to [42] and [37]. On the one hand [42, p. 63ff.] covers the $\mathcal{H}_2$ inner-product of transfer functions in the ODE-case which is then used to prove $\mathcal{H}_2$ pseudo-optimality. On the other hand controllability and observability Gramians for DAEs (presented in Definition 2.32) are analyzed in the context of generalized Lyapunov equations in [37]. The main result of this section is in some sense a generalization, connecting both areas.

In order to formulate the $\mathcal{H}_2$ inner-product of the DAE-systems $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ and $(\mathbf{E}_\mathrm{H},\, \mathbf{A}_\mathrm{H},\, \mathbf{B}_\mathrm{H},\, \mathbf{C}_\mathrm{H},\, \mathbf{x}_{\mathrm{H},0})$ of dimension n and n_H, their transfer functions $\mathbf{G}(s)$ and $\mathbf{G}_\mathrm{H}(s)$ must

- be asymptotically stable and strictly proper (i. e. $\mathbf{G}(s),\, \mathbf{G}_\mathrm{H}(s) \in \mathcal{H}_2^{(p,m)}$),
- have the same count of inputs and outputs (i. e. $p = p_\mathrm{H}$ and $m = m_\mathrm{H}$) and
- *allow* realizations with real-valued system matrices.[1]

As it is assumed that $\mathbf{G}(s)$ and $\mathbf{G}_\mathrm{H}(s)$ have the same dimension, the superscript $(p,\, m)$ will be omitted in order to shorten the notation. Furthermore keep in mind, that the realizations $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ and $[\mathbf{E}_\mathrm{H},\, \mathbf{A}_\mathrm{H},\, \mathbf{B}_\mathrm{H},\, \mathbf{C}_\mathrm{H}]$ do not have to be real-valued themselves.

From (4.2) it follows, that

$$
\begin{aligned}
\langle \mathbf{G},\, \mathbf{G}_\mathrm{H} \rangle_{\mathcal{H}_2} &= \int_0^\infty \mathrm{tr}\left\{ \mathbf{G}_\mathrm{H}(t)\, \mathbf{G}^*(t) \right\} \mathrm{d}t \overset{(4.4)}{=} \int_0^\infty \mathrm{tr}\left\{ \mathbf{G}(t)\, \mathbf{G}_\mathrm{H}^*(t) \right\} \mathrm{d}t \\
&= \mathrm{tr}\left\{ \int_0^\infty \mathbf{G}(t)\, \mathbf{G}_\mathrm{H}^*(t) \mathrm{d}t \right\} \\
&\overset{(2.38)}{=} \mathrm{tr}\left\{ \int_0^\infty \left(\tilde{\mathbf{C}}\, e^{\tilde{\mathbf{E}}^\mathrm{D}\, \tilde{\mathbf{A}}\, t}\, \tilde{\mathbf{E}}^\mathrm{D}\, \tilde{\mathbf{B}}\, \tilde{\mathbf{B}}_\mathrm{H}^*\, \tilde{\mathbf{E}}_\mathrm{H}^{\mathrm{D}*}\, e^{\tilde{\mathbf{A}}_\mathrm{H}^*\, \tilde{\mathbf{E}}_\mathrm{H}^{\mathrm{D}*}\, t}\, \tilde{\mathbf{C}}_\mathrm{H}^* \right) \mathrm{d}t \right\} \,,
\end{aligned}
\tag{4.6}
$$

where $[\tilde{\mathbf{E}},\, \tilde{\mathbf{A}},\, \tilde{\mathbf{B}},\, \tilde{\mathbf{C}}]$ and $[\tilde{\mathbf{E}}_\mathrm{H},\, \tilde{\mathbf{A}}_\mathrm{H},\, \tilde{\mathbf{B}}_\mathrm{H},\, \tilde{\mathbf{C}}_\mathrm{H}]$ are realizations of $\mathbf{G}(s)$ and $\mathbf{G}_\mathrm{H}(s)$ in Weierstraß canonical form obtained through transformations with the regular matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{P}_\mathrm{H}$, $\mathbf{Q}_\mathrm{H}$ according to Lemma 2.5.

Using (2.41) one can show, that the integral still converges if the terms $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{C}}_\mathrm{H}^*$ are extracted to the left and right:

$$
\langle \mathbf{G},\, \mathbf{G}_\mathrm{H} \rangle_{\mathcal{H}_2} = \mathrm{tr}\left\{ \tilde{\mathbf{C}} \int_0^\infty \left(e^{\tilde{\mathbf{E}}^\mathrm{D}\, \tilde{\mathbf{A}}\, t}\, \tilde{\mathbf{E}}^\mathrm{D}\, \tilde{\mathbf{B}}\, \tilde{\mathbf{B}}_\mathrm{H}^*\, \tilde{\mathbf{E}}_\mathrm{H}^{\mathrm{D}*}\, e^{\tilde{\mathbf{A}}_\mathrm{H}^*\, \tilde{\mathbf{E}}_\mathrm{H}^{\mathrm{D}*}\, t} \right) \mathrm{d}t\, \tilde{\mathbf{C}}_\mathrm{H}^* \right\} \,. \tag{4.7}
$$

A similar strategy as in the proof of Theorem 2.26 helps to simplify (4.7) to

$$
\langle \mathbf{G},\, \mathbf{G}_\mathrm{H} \rangle_{\mathcal{H}_2} = \mathrm{tr}\left\{ \tilde{\mathbf{C}} \begin{bmatrix} \int_0^\infty \left(e^{\mathbf{J}\, t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_\mathrm{H}^*\, t} \right) \mathrm{d}t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tilde{\mathbf{C}}_\mathrm{H}^* \right\} \,, \tag{4.8}
$$

or alternatively

$$
\langle \mathbf{G},\, \mathbf{G}_\mathrm{H} \rangle_{\mathcal{H}_2} = \mathrm{tr}\left\{ \tilde{\mathbf{C}}\, \tilde{\mathbf{X}}\, \tilde{\mathbf{C}}_\mathrm{H}^* \right\} \,, \tag{4.9}
$$

[1] This assumption is necessary for commutativity of the $\mathcal{H}_2$ inner-product according to Lemma 4.7.

with the new unknown but constant matrix

$$\tilde{\mathbf{X}} := \begin{bmatrix} \tilde{\mathbf{X}}_{ff} & \tilde{\mathbf{X}}_{f\infty} \\ \tilde{\mathbf{X}}_{\infty f} & \tilde{\mathbf{X}}_{\infty\infty} \end{bmatrix} = \begin{bmatrix} \int_0^\infty \left(e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t} \right) \mathrm{d}t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{C}^{n \times n_{\mathrm{H}}} \,. \tag{4.10}$$

The following result shows, that the computation of the integral expression in (4.10) can be avoided by solving a generalized Sylvester equation (with additional constraint).

Lemma 4.8. *Let* $[\tilde{\mathbf{E}}, \tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}]$ *and* $[\tilde{\mathbf{E}}_{\mathrm{H}}, \tilde{\mathbf{A}}_{\mathrm{H}}, \tilde{\mathbf{B}}_{\mathrm{H}}, \tilde{\mathbf{C}}_{\mathrm{H}}]$ *be realizations of two asymptotically stable and strictly proper DAE-systems in Weierstraß canonical form, which have the same count of input- and output-variables, i. e.* $p = p_{\mathrm{H}}$ *and* $m = m_{\mathrm{H}}$*, and allow realizations with real-valued system matrices.*

Then the constant matrix $\tilde{\mathbf{X}}$*, as defined in (4.10), is the unique solution of the generalized projected Sylvester equation*

$$\tilde{\mathbf{A}}\,\tilde{\mathbf{X}}\,\tilde{\mathbf{E}}_{\mathrm{H}}^* + \tilde{\mathbf{E}}\,\tilde{\mathbf{X}}\,\tilde{\mathbf{A}}_{\mathrm{H}}^* + \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tilde{\mathbf{B}}\,\tilde{\mathbf{B}}_{\mathrm{H}}^* \begin{bmatrix} \mathbf{I}_{n_{\mathrm{H}f}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{0} \tag{4.11}$$

whilst taking into account one of the three constraints:

$$\begin{aligned} &\textit{(i)}\ \tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tilde{\mathbf{X}}\,, \\ &\textit{(ii)}\ \tilde{\mathbf{X}} = \tilde{\mathbf{X}} \begin{bmatrix} \mathbf{I}_{n_{\mathrm{H}f}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \\ &\textit{(iii)}\ \tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tilde{\mathbf{X}} \begin{bmatrix} \mathbf{I}_{n_{\mathrm{H}f}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \end{aligned} \tag{4.12}$$

Proof. Using the special structure of the system matrices in Weierstraß canonical form allows to decompose (4.11) into four decoupled equations

$$\mathbf{J}\,\tilde{\mathbf{X}}_{ff} + \tilde{\mathbf{X}}_{ff}\,\mathbf{J}_{\mathrm{H}}^* + \tilde{\mathbf{B}}_f\,\tilde{\mathbf{B}}_{\mathrm{H}f}^* = \mathbf{0}\,, \tag{4.13}$$

$$\tilde{\mathbf{X}}_{\infty f} + \mathbf{N}\,\tilde{\mathbf{X}}_{\infty f}\,\mathbf{J}_{\mathrm{H}}^* = \mathbf{0}\,, \tag{4.14}$$

$$\mathbf{J}\,\tilde{\mathbf{X}}_{f\infty}\,\mathbf{N}_{\mathrm{H}}^* + \tilde{\mathbf{X}}_{f\infty} = \mathbf{0}\,, \tag{4.15}$$

$$\tilde{\mathbf{X}}_{\infty\infty}\,\mathbf{N}_{\mathrm{H}}^* + \mathbf{N}\,\tilde{\mathbf{X}}_{\infty\infty} = \mathbf{0}\,. \tag{4.16}$$

Inserting $\tilde{\mathbf{X}}_{ff} = \int_0^\infty \left(e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t} \right) \mathrm{d}t$ into (4.13) shows, that the $\tilde{\mathbf{X}}_{ff}$-part of (4.11) matches the desired solution from (4.10):

$$\begin{aligned} \mathbf{J} \int_0^\infty \left(e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t} \right) \mathrm{d}t + \int_0^\infty \left(e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t} \right) \mathrm{d}t\, \mathbf{J}_{\mathrm{H}}^* + \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^* &= \\ = \int_0^\infty \left(\mathbf{J}\, e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t} + e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t}\, \mathbf{J}_{\mathrm{H}}^* \right) \mathrm{d}t + \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^* & \\ = \int_0^\infty \left(\frac{\mathrm{d}}{\mathrm{d}t} \left[e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t} \right] \right) \mathrm{d}t + \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^* & \\ = e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t} \Big|_0^\infty + \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^* & \\ = \underbrace{\lim_{t\to\infty} e^{\mathbf{J}\,t}\, \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^*\, e^{\mathbf{J}_{\mathrm{H}}^*\,t}}_{\text{asymptotically stable} \Rightarrow \mathbf{0}} - \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^* + \tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_{\mathrm{H}f}^* = \mathbf{0}\,. & \end{aligned} \tag{4.17}$$

Applying Theorem 3.13 to (4.14) and (4.15) leads to the conclusion, that the solutions $\tilde{\mathbf{X}}_{\infty f}$ and $\tilde{\mathbf{X}}_{f\infty}$ are unique (see Table 4.1). Since $\tilde{\mathbf{X}}_{\infty f} = \mathbf{0}$ and $\tilde{\mathbf{X}}_{f\infty} = \mathbf{0}$ are the trivial solutions of (4.14) and (4.15), also the $\tilde{\mathbf{X}}_{\infty f}$- and $\tilde{\mathbf{X}}_{f\infty}$-parts of (4.11) are equivalent to (4.10).

Table 4.1: Analysis of Equations (4.13) to (4.16) regarding existence and uniqueness of the solutions according to Theorem 3.13.

Eq.	$\triangleq \mathbf{A}$	$\triangleq \mathbf{B}$	$\triangleq \mathbf{C}$	$\triangleq \mathbf{D}$	$(\lambda\,\mathbf{C} - \mathbf{A})$	$(\lambda\,\mathbf{B} - \mathbf{D})$	$\lambda(\mathbf{C},\,\mathbf{A})$	$\lambda(\mathbf{B},\,\mathbf{D})$
(4.13)	$\mathbf{J}$	$\mathbf{I}_{n_{\mathrm{H}f}}$	$-\mathbf{I}_{n_f}$	$\mathbf{J}_{\mathrm{H}}^*$	regular	regular	Re > 0	Re < 0
(4.14)	$\mathbf{I}_{n_\infty}$	$\mathbf{I}_{n_{\mathrm{H}f}}$	$-\mathbf{N}$	$\mathbf{J}_{\mathrm{H}}^*$	regular	regular	$\pm\infty$	finite
(4.15)	$\mathbf{J}$	$\mathbf{N}_{\mathrm{H}}^*$	$-\mathbf{I}_{n_f}$	$\mathbf{I}_{n_{\mathrm{H}\infty}}$	regular	regular	finite	$\pm\infty$
(4.16)	$\mathbf{I}_{n_\infty}$	$\mathbf{N}_{\mathrm{H}}^*$	$-\mathbf{N}$	$\mathbf{I}_{n_{\mathrm{H}\infty}}$	regular	regular	$\pm\infty$	$\pm\infty$

It is left to prove, that $\tilde{\mathbf{X}}_{\infty\infty} = \mathbf{0}$ is the unique solution of (4.16). According to Theorem 3.13 there are infinitely many solutions of (4.16), $\tilde{\mathbf{X}}_{\infty\infty} = \mathbf{0}$ is only one of them. Through the demand, that $\tilde{\mathbf{X}}$ fulfils one of the constraints in (4.12) it is guaranteed, that $\tilde{\mathbf{X}}_{\infty\infty} = \mathbf{0}$ is the only solution of (4.16) which completes the proof. ■

Since the Weierstraß canonical form is not known in the general case, a more convenient form of Lemma 4.8 is needed, which can be obtained by back transformation with

$$\begin{aligned} &\tilde{\mathbf{E}} = \mathbf{P}\,\mathbf{E}\,\mathbf{Q}\,, && \tilde{\mathbf{A}} = \mathbf{P}\,\mathbf{A}\,\mathbf{Q}\,, && \tilde{\mathbf{B}} = \mathbf{P}\,\mathbf{B}\,, && \tilde{\mathbf{C}} = \mathbf{C}\,\mathbf{Q}\,, \\ &\tilde{\mathbf{E}}_{\mathrm{H}}^* = \mathbf{Q}_{\mathrm{H}}^*\,\mathbf{E}_{\mathrm{H}}^*\,\mathbf{P}_{\mathrm{H}}^*\,, && \tilde{\mathbf{A}}_{\mathrm{H}}^* = \mathbf{Q}_{\mathrm{H}}^*\,\mathbf{A}_{\mathrm{H}}^*\,\mathbf{P}_{\mathrm{H}}^*\,, && \tilde{\mathbf{B}}_{\mathrm{H}}^* = \mathbf{B}_{\mathrm{H}}^*\,\mathbf{P}_{\mathrm{H}}^*\,, && \tilde{\mathbf{C}}_{\mathrm{H}}^* = \mathbf{Q}_{\mathrm{H}}^*\,\mathbf{C}_{\mathrm{H}}^* \end{aligned} \tag{4.18}$$

and

$$\mathbf{X} := \mathbf{Q}\,\tilde{\mathbf{X}}\,\mathbf{Q}_{\mathrm{H}}^*\,. \tag{4.19}$$

This is summarized in Theorem 4.9, which represents the main result of this section:

Key Theorem 4.9. *Let*

- $\mathbf{G}(s)$ *and* $\mathbf{G}_{\mathrm{H}}(s)$ *be transfer functions of asymptotically stable and strictly proper DAE-systems with the same count of input- and output variables, i. e.* $p = p_{\mathrm{H}}$ *and* $m = m_{\mathrm{H}}$*, and which allow realizations with real-valued system matrices,*
- $[\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C}]$ *and* $[\mathbf{E}_{\mathrm{H}},\,\mathbf{A}_{\mathrm{H}},\,\mathbf{B}_{\mathrm{H}},\,\mathbf{C}_{\mathrm{H}}]$ *be realizations of* $\mathbf{G}(s)$ *and* $\mathbf{G}_{\mathrm{H}}(s)$*,*
- $\mathbf{\Pi}_l^f$*,* $\mathbf{\Pi}_r^f$ *and* $\mathbf{\Pi}_{\mathrm{H}l}^f$*,* $\mathbf{\Pi}_{\mathrm{H}r}^f$ *be the spectral projectors related to* $\lambda\,\mathbf{E} - \mathbf{A}$ *and* $\lambda\,\mathbf{E}_{\mathrm{H}} - \mathbf{A}_{\mathrm{H}}$ *respectively according to Definition 2.8,*

Then the $\mathcal{H}_2$ *inner-product is given by*

$$\langle \mathbf{G},\,\mathbf{G}_{\mathrm{H}}\rangle_{\mathcal{H}_2} = \mathrm{tr}\,(\mathbf{C}\,\mathbf{X}\,\mathbf{C}_{\mathrm{H}}^*) = \mathrm{tr}\,(\mathbf{B}^*\,\mathbf{Y}\,\mathbf{B}_{\mathrm{H}}) \tag{4.20}$$

with $\mathbf{X}$ *and* $\mathbf{Y}$ *as the* unique *solutions of*

$$\begin{gathered} \mathbf{A}\,\mathbf{X}\,\mathbf{E}_{\mathrm{H}}^* + \mathbf{E}\,\mathbf{X}\,\mathbf{A}_{\mathrm{H}}^* + \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{B}_{\mathrm{H}}^*\,\mathbf{\Pi}_{\mathrm{H}l}^{f*} = \mathbf{0} \\ \textit{with } \mathbf{X} = \mathbf{\Pi}_r^f\,\mathbf{X} \textit{ or } \mathbf{X} = \mathbf{X}\,\mathbf{\Pi}_{\mathrm{H}r}^{f*} \textit{ or } \mathbf{X} = \mathbf{\Pi}_r^f\,\mathbf{X}\,\mathbf{\Pi}_{\mathrm{H}r}^{f*}\,, \end{gathered} \tag{4.21}$$

$$\begin{gathered} \mathbf{A}^*\,\mathbf{Y}\,\mathbf{E}_{\mathrm{H}} + \mathbf{E}^*\,\mathbf{Y}\,\mathbf{A}_{\mathrm{H}} + \mathbf{\Pi}_r^{f*}\,\mathbf{C}^*\,\mathbf{C}_{\mathrm{H}}\,\mathbf{\Pi}_{\mathrm{H}r}^f = \mathbf{0} \\ \textit{with } \mathbf{Y} = \mathbf{\Pi}_l^{f*}\,\mathbf{Y} \textit{ or } \mathbf{Y} = \mathbf{Y}\,\mathbf{\Pi}_{\mathrm{H}l}^f \textit{ or } \mathbf{Y} = \mathbf{\Pi}_l^{f*}\,\mathbf{Y}\,\mathbf{\Pi}_{\mathrm{H}l}^f\,. \end{gathered} \tag{4.22}$$

Proof. The relation for $\mathbf{X}$ in (4.21) directly follows from a back transformation of Lemma 4.8 and (4.9) according to (4.18). Due to the duality principle in linear systems, a similar relationship for $\mathbf{Y}$ (4.22) holds. ∎

Theorem 4.9 describes the general case of connecting two DAEs. For the purposes of $\mathcal{H}_2$ pseudo-optimal reduction described in the following section, the combination of a DAE ($\mathbf{G}(s)$) and an ODE ($\mathbf{G}_{\mathrm{H}}(s)$) is considered, which helps to simplify the computation of $\mathbf{X}$ and $\mathbf{Y}$:

Corollary 4.10. *Let all conditions of Theorem 4.9 hold. If additionally* $\det(\mathbf{E}_{\mathrm{H}}) \neq 0$, *then* $\mathbf{X}$ *and* $\mathbf{Y}$ *are the* unique *solutions of*

$$\mathbf{A}\,\mathbf{X}\,\mathbf{E}_{\mathrm{H}}^{*} + \mathbf{E}\,\mathbf{X}\,\mathbf{A}_{\mathrm{H}}^{*} + \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{B}_{\mathrm{H}}^{*} = \mathbf{0}\,, \tag{4.23}$$

$$\mathbf{A}^{*}\,\mathbf{Y}\,\mathbf{E}_{\mathrm{H}} + \mathbf{E}^{*}\,\mathbf{Y}\,\mathbf{A}_{\mathrm{H}} + \mathbf{\Pi}_r^{f*}\,\mathbf{C}^{*}\,\mathbf{C}_{\mathrm{H}} = \mathbf{0}\,. \tag{4.24}$$

Proof. In order the verify (4.23) and (4.24) note that $\det(\mathbf{E}_{\mathrm{H}}) \neq 0$ implies $\mathbf{\Pi}_{\mathrm{H}l}^f = \mathbf{\Pi}_{\mathrm{H}r}^f = \mathbf{I}_{n_{\mathrm{H}}}$. Therefore the generalized projected Sylvester equations have unique solutions, which is why the additional constraints are not needed anymore. ∎

It is worth noting, that Theorem 4.9 describes a generalization of the proper controllability and observability Gramians of DAE-systems (see Definition 2.32):

Corollary 4.11. *Let all conditions of Theorem 4.9 hold and additionally* $\mathbf{E}_{\mathrm{H}} = \mathbf{E}$, $\mathbf{A}_{\mathrm{H}} = \mathbf{A}$, $\mathbf{B}_{\mathrm{H}} = \mathbf{B}$, $\mathbf{C}_{\mathrm{H}} = \mathbf{C}$ *i. e.* $\mathbf{G}_{\mathrm{H}}(s) = \mathbf{G}(s)$.

Then

(i) the matrices $\mathbf{X}$ *and* $\mathbf{Y}$ *coincide with the proper controllability and observability Gramians* $\mathbf{\Gamma}^{\mathrm{pc}}$ *and* $\mathbf{\Gamma}^{\mathrm{po}}$ *respectively and*

(ii) the $\mathcal{H}_2$ *norm of* $\mathbf{G}(s)$ *reads as*

$$\|\mathbf{G}\|_{\mathcal{H}_2}^2 = \mathrm{tr}(\mathbf{C}\,\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{C}^{*}) = \mathrm{tr}(\mathbf{B}^{*}\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{B})\,. \tag{4.25}$$

Proof. The first part follows from a comparison of (4.21) and (4.22) with (2.47) and (2.48). The second part is a consequence of (4.20) and (4.3) and matches perfectly with the results in [40, p. 845] (therein denoted by the strictly proper part $\|\mathbf{G}_{sp}\|_{\mathbb{H}_2}^2$). ∎

4.2 $\mathcal{H}_2$ Pseudo-Optimal Reduction of DAEs

This section is strongly related to [42], which introduces the PORK algorithm as $\mathcal{H}_2$ pseudo-optimal reduction scheme for ODE-systems. In the following these results are revised and extended to the DAE-case. Note that [42] uses a different notation of rational Krylov subspaces. Furthermore the derivation is limited to special use cases (regarding the interpolation data) in order to keep things simple. In contrast the following investigations represent a complete proof of the most general case.

For this purpose three different realizations of the reduced transfer function $\mathbf{G}_{\mathrm{r}}(s)$ are considered (see Figure 4.1):

- $[\mathbf{E}_{\mathrm{M}}, \mathbf{A}_{\mathrm{M}}, \mathbf{B}_{\mathrm{M}}, \mathbf{C}_{\mathrm{M}}]$ related to the system $\mathbf{\Sigma}_{\mathrm{M}} := (\mathbf{E}_{\mathrm{M}}, \mathbf{A}_{\mathrm{M}}, \mathbf{B}_{\mathrm{M}}, \mathbf{C}_{\mathrm{M}}, \mathbf{x}_{\mathrm{M},0})$,

- $[\mathbf{E}_\mathrm{F},\ \mathbf{A}_\mathrm{F},\ \mathbf{B}_\mathrm{F},\ \mathbf{C}_\mathrm{F}]$ related to the system $\mathbf{\Sigma}_\mathrm{F} := (\mathbf{E}_\mathrm{F},\ \mathbf{A}_\mathrm{F},\ \mathbf{B}_\mathrm{F},\ \mathbf{C}_\mathrm{F},\ \mathbf{x}_{\mathrm{F},0})$ and
- $[\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r}]$ related to the system $\mathbf{\Sigma}_\mathrm{r} := (\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r},\ \mathbf{x}_{\mathrm{r},0})$.

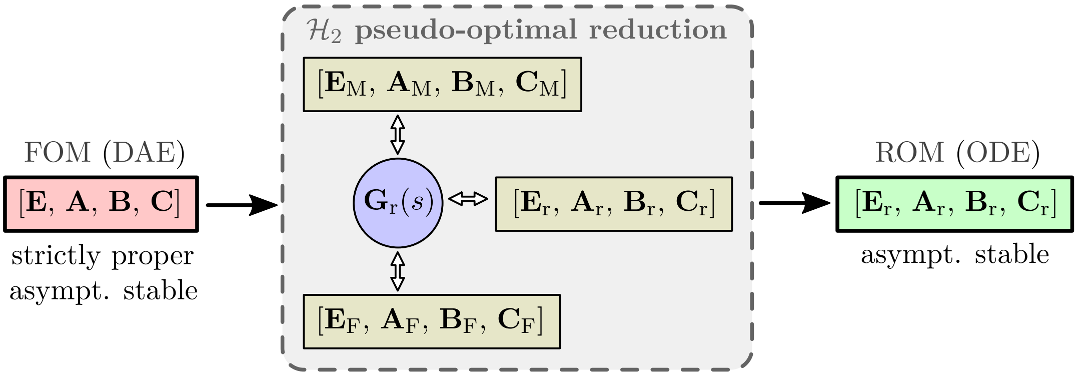

$\mathcal{H}_2$ pseudo-optimal reduction
$[\mathbf{E}_\mathrm{M},\ \mathbf{A}_\mathrm{M},\ \mathbf{B}_\mathrm{M},\ \mathbf{C}_\mathrm{M}]$
FOM (DAE)
$[\mathbf{E},\ \mathbf{A},\ \mathbf{B},\ \mathbf{C}]$
strictly proper
asympt. stable
$\mathbf{G}_\mathrm{r}(s)$
$[\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r}]$
ROM (ODE)
$[\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r}]$
asympt. stable
$[\mathbf{E}_\mathrm{F},\ \mathbf{A}_\mathrm{F},\ \mathbf{B}_\mathrm{F},\ \mathbf{C}_\mathrm{F}]$

Figure 4.1: Realizations of the ROM during $\mathcal{H}_2$ pseudo-optimal reduction: while $[\mathbf{E}_\mathrm{M},\ \mathbf{A}_\mathrm{M},\ \mathbf{B}_\mathrm{M},\ \mathbf{C}_\mathrm{M}]$ and $[\mathbf{E}_\mathrm{F},\ \mathbf{A}_\mathrm{F},\ \mathbf{B}_\mathrm{F},\ \mathbf{C}_\mathrm{F}]$ are necessary to derive the PORK algorithm, $[\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r}]$ denotes the actual result of the reduction process.

Since they share the same transfer function, $\mathbf{\Sigma}_\mathrm{M}$, $\mathbf{\Sigma}_\mathrm{F}$ and $\mathbf{\Sigma}_\mathrm{r}$ are restricted system equivalent (see Definition 2.19). In view of the proof of $\mathcal{H}_2$ pseudo-optimality they are handled separately at first. Their connection (and especially the equality $\mathbf{G}_\mathrm{M}(s) = \mathbf{G}_\mathrm{F}(s) = \mathbf{G}_\mathrm{r}(s)$) will be shown later on.

In contrast to $\mathbf{\Sigma}_\mathrm{r}$, which denotes the final result of the reduction process, $\mathbf{\Sigma}_\mathrm{M}$ and $\mathbf{\Sigma}_\mathrm{F}$ are only of theoretical interest, i. e. they are not needed for implementation. Note that although $\mathbf{\Sigma}_\mathrm{M}$ and $\mathbf{\Sigma}_\mathrm{F}$ are of special structure, there are no additional constraints on the shape of the FOM (asymptotic stability and strictly properness are sufficient).

Because the DAE-system describing the FOM is assumed to be strictly proper, it is demanded without loss of generality, that the ROM is of ODE-type, thus $\det(\mathbf{E}_\mathrm{M}) \neq 0$, $\det(\mathbf{E}_\mathrm{F}) \neq 0$ and $\det(\mathbf{E}_\mathrm{r}) \neq 0$. This way additional (unnecessary) algebraic equations are avoided.

4.2.1 The Mirrored Jordan Canonical Form

In order to shorten the following proofs, the naming of a special structure of a matrix is introduced. For this purpose *mirrored Jordan blocks* are defined:

Definition 4.12. Let $\mathbf{Z} \in \mathbb{C}^{u\times u}$ be structured as

$$\mathbf{Z} = \begin{bmatrix} \lambda & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & \lambda \end{bmatrix}, \tag{4.26}$$

with $\lambda \in \mathbb{C}$. Then $\mathbf{Z}$ is called to be a *mirrored Jordan block.*

In contrast to “regular” Jordan blocks, the structure is “mirrored” around the diagonal while switching the sign of the minor diagonal. Similar to the assembly of Jordan block in the Jordan canonical form, mirrored Jordan blocks can be concatenated to the *mirrored Jordan canonical form*:

Definition 4.13. Let $\mathbf{X} \in \mathbb{C}^{q\times q}$ be structured as

$$\mathbf{X} = \operatorname{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_i, \ldots, \mathbf{X}_s)\ , \tag{4.27a}$$

$$\mathbf{X}_i = \operatorname{diag}(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{ij}, \ldots, \mathbf{X}_{ir_i}) \qquad \forall\, i = 1, \ldots, s\ , \tag{4.27b}$$

$$\mathbf{X}_{ij} = \begin{bmatrix} \lambda_i & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & \lambda_i \end{bmatrix} \in \mathbb{C}^{q_{ij}\times q_{ij}} \qquad \forall\, j = 1, \ldots, r_i\ , \tag{4.27c}$$

with pairwise different eigenvalues $\lambda_i \in \mathbb{C}$, i. e. $\lambda_i \neq \lambda_w$ for $i \neq w$, and the mirrored Jordan blocks $\mathbf{X}_{ij}$. Then $\mathbf{X}$ is called to be in *mirrored Jordan canonical form.*

Note that the structures of $\mathbf{X}$ in Definition 4.13 and $\mathbf{S}_V^{\mathrm{P}}$ (which is in Jordan canonical form, see (3.29)) are very similar. In the following both matrices will be connected as part of the proof of $\mathcal{H}_2$ pseudo-optimality. Therefore the same dimension $(q \times q)$, segmentation and indexing (e. g. $i = 1, \ldots, s$) is used.

Analogous to the Jordan canonical form, every (quadratic) matrix can be transformed into mirrored Jordan canonical form:

Lemma 4.14. *For every* $\mathbf{Y} \in \mathbb{C}^{q\times q}$ *there exists a regular transformation matrix* $\mathbf{T} \in \mathbb{C}^{q\times q}$*, such that* $\mathbf{X} = \mathbf{T}^{-1}\,\mathbf{Y}\,\mathbf{T}$ *is in mirrored Jordan canonical form.*

Proof. Since $\mathbf{Y}$ is quadratic, there exists a regular transformation matrix $\mathbf{T}_J \in \mathbb{C}^{q\times q}$ such that $\mathbf{J} = \mathbf{T}_J^{-1}\,\mathbf{Y}\,\mathbf{T}_J$ is in Jordan canonical form [46, p. 610]. Now consider one Jordan block $\mathbf{J}_{ij} \in \mathbb{C}^{q_{ij}\times q_{ij}}$ of $\mathbf{J}$ and its transformation with $\mathbf{T}_\downarrow \in \mathbb{C}^{q_{ij}\times q_{ij}}$ (switching to lower triangular form[2]) and $\mathbf{T}_N \in \mathbb{C}^{q_{ij}\times q_{ij}}$ (negation of ones):

$$\mathbf{X}_{ij} = \underbrace{\begin{bmatrix} -1 & & & \\ & 1 & & \\ & & -1 & \\ & & & \ddots \end{bmatrix}}_{\mathbf{T}_N^{-1}} \underbrace{\begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}}_{\mathbf{T}_\downarrow^{-1}} \underbrace{\begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}}_{\mathbf{J}_{ij}} \underbrace{\begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}}_{\mathbf{T}_\downarrow} \underbrace{\begin{bmatrix} -1 & & & \\ & 1 & & \\ & & -1 & \\ & & & \ddots \end{bmatrix}}_{\mathbf{T}_N} \tag{4.28}$$

The matrix $\mathbf{X}_{ij}$ is one of the desired mirrored Jordan blocks of $\mathbf{X}$, $\mathbf{T}_\downarrow$ is anti-diagonal and $\mathbf{T}_N$ is diagonal with alternating signs, i. e. $T_{N,ww} = (-1)^w$. Finally the overall transformation matrix $\mathbf{T}$ can be assembled through

$$\mathbf{T} = \mathbf{T}_J\ \operatorname{diag}(\mathbf{T}_\downarrow, \ldots, \mathbf{T}_\downarrow)\ \operatorname{diag}(\mathbf{T}_N, \ldots, \mathbf{T}_N)\ , \tag{4.29}$$

where $\mathbf{T}_\downarrow$ and $\mathbf{T}_N$ adapt their dimensions according to $\mathbf{J}_{ij}$. ■

Using Definition 4.13, correspondingly structured realizations of ODE-systems are introduced:

Definition 4.15. A realization $[\mathbf{E}^{\reflectbox{J}}, \mathbf{A}^{\reflectbox{J}}, \mathbf{B}^{\reflectbox{J}}, \mathbf{C}^{\reflectbox{J}}]$ of an ODE-system is called to be in *mirrored Jordan canonical form*, if $\mathbf{E}^{\reflectbox{J}} = \mathbf{I}_q$ and $\mathbf{A}^{\reflectbox{J}}$ is in mirrored Jordan canonical form as described in Definition 4.13. Note the superscript $\reflectbox{J}$ which is an indicator for the mirrored Jordan canonical form. ODE

[2] Since all diagonal elements of $\mathbf{J}_{ij}$ are equal, the transformation with $\mathbf{T}_\downarrow$ and $\mathbf{T}_\downarrow^{-1}$ corresponds to a true flipping over the diagonal.

An adaption of Definition 4.15 to the more general case of DAE-systems using a modified Weierstraß canonical form is possible. Since the following proofs would not benefit from such a notation, it will not be defined. Instead a restriction to strictly proper DAEs is made, which allows to use results of ODE-theory as an intermediate step:

Lemma 4.16. *For every strictly proper DAE-system, there exists a realization of the transfer function* $\mathbf{G}(s)$ *in mirrored Jordan canonical form.*

Proof. Since the system is strictly proper, there exists a realization $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ of $\mathbf{G}(s)$ in ODE-form (see Corollary 2.31), i. e. $\det(\mathbf{E}) \neq 0$, such that

$$\mathbf{G}(s) = \mathbf{C}\,(s\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{B} = \mathbf{C}\left(s\,\mathbf{I}_q - \mathbf{E}^{-1}\,\mathbf{A}\right)^{-1}\,\mathbf{E}^{-1}\,\mathbf{B}\;, \tag{4.30}$$

or rather $\mathbf{G}(s) = \mathbf{C}\,(s\,\mathbf{I}_q - \mathbf{Y})^{-1}\,\mathbf{E}^{-1}\,\mathbf{B}$ with $\mathbf{Y} = \mathbf{E}^{-1}\,\mathbf{A}$. According to Lemma 4.14 there exists a regular transformation matrix $\mathbf{T} \in \mathbb{C}^{q\times q}$ such that $\mathbf{X} = \mathbf{T}^{-1}\,\mathbf{Y}\,\mathbf{T}$ is in mirrored Jordan canonical form. Therefore one can find

$$\mathbf{G}(s) = \mathbf{C}\left(s\,\mathbf{I}_q - \mathbf{T}\,\mathbf{X}\,\mathbf{T}^{-1}\right)^{-1}\,\mathbf{E}^{-1}\,\mathbf{B} = \mathbf{C}\,\mathbf{T}\,(s\,\mathbf{I}_q - \mathbf{X})^{-1}\,\mathbf{T}^{-1}\,\mathbf{E}^{-1}\,\mathbf{B}\;, \tag{4.31}$$

and finally the realization $[\mathbf{E}^{\mathfrak{l}},\, \mathbf{A}^{\mathfrak{l}},\, \mathbf{B}^{\mathfrak{l}},\, \mathbf{C}^{\mathfrak{l}}]$ with

$$\mathbf{E}^{\mathfrak{l}} = \mathbf{I}_q\;, \qquad \mathbf{A}^{\mathfrak{l}} = \mathbf{T}^{-1}\,\mathbf{E}^{-1}\,\mathbf{A}\,\mathbf{T}\;, \qquad \mathbf{B}^{\mathfrak{l}} = \mathbf{T}^{-1}\,\mathbf{E}^{-1}\,\mathbf{B}\;, \qquad \mathbf{C}^{\mathfrak{l}} = \mathbf{C}\,\mathbf{T}\;, \tag{4.32}$$

which is in mirrored Jordan canonical form. ■

The following theorem is an extension of two special cases discussed in [42, p. 66ff.] to the most general form. It presents a convenient formulation of the $\mathcal{H}_2$ inner-product of $\mathbf{G}(s)$ (FOM) and $\mathbf{G}_{\mathrm{M}}(s)$ (ROM, corresponds to $\mathbf{\Sigma}_{\mathrm{M}}$) via the moments of $\mathbf{G}(s)$. For this purpose a realization of $\mathbf{G}_{\mathrm{M}}(s)$ in mirrored Jordan canonical form is used, whose structure is exploited. Although not immediatly obvious, Theorem 4.17 can be considered as the key to the proof of the PORK algorithm for DAEs, since it makes use of the main result of the last section (i. e. Corollary 4.10 as a special case of Theorem 4.9).

Key Theorem 4.17. *Let* $\mathbf{G}(s)$ *be the transfer function of an asymptotically stable and strictly proper DAE-system and let* $\mathbf{G}_{\mathrm{M}}(s)$ *be the transfer function of an asymptotically stable ODE-system. Let* $\mathbf{G}(s)$ *and* $\mathbf{G}_{\mathrm{M}}(s)$ *have the same count of input- and output-variables, i. e.* $p = p_{\mathrm{M}}$ *and* $m = m_{\mathrm{M}}$*, and allow realizations with real-valued system matrices. Let* $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ *and* $[\mathbf{E}^{\mathfrak{l}}_{\mathrm{M}},\, \mathbf{A}^{\mathfrak{l}}_{\mathrm{M}},\, \mathbf{B}^{\mathfrak{l}}_{\mathrm{M}},\, \mathbf{C}^{\mathfrak{l}}_{\mathrm{M}}]$ *be realizations of* $\mathbf{G}(s)$ *and* $\mathbf{G}_{\mathrm{M}}(s)$ *respectively where* $[\mathbf{E}^{\mathfrak{l}}_{\mathrm{M}},\, \mathbf{A}^{\mathfrak{l}}_{\mathrm{M}},\, \mathbf{B}^{\mathfrak{l}}_{\mathrm{M}},\, \mathbf{C}^{\mathfrak{l}}_{\mathrm{M}}]$ *is in mirrored Jordan canonical form according to Definition 4.15.*

Consider the decomposition

$$\begin{aligned}
\mathbf{A}^{\mathfrak{l}}_{\mathrm{M}} &= \operatorname{diag}(\mathbf{A}^{\mathfrak{l}}_{\mathrm{M}1},\, \dots,\, \mathbf{A}^{\mathfrak{l}}_{\mathrm{M}i},\, \dots,\, \mathbf{A}^{\mathfrak{l}}_{\mathrm{M}s})\;,\\
\mathbf{B}^{\mathfrak{l}*}_{\mathrm{H}} &= \left[\mathbf{B}^{\mathfrak{l}*}_{\mathrm{M}1},\, \dots,\, \mathbf{B}^{\mathfrak{l}*}_{\mathrm{M}i},\, \dots,\, \mathbf{B}^{\mathfrak{l}*}_{\mathrm{M}s}\right]\;,\\
\mathbf{C}^{\mathfrak{l}}_{\mathrm{H}} &= \left[\mathbf{C}^{\mathfrak{l}}_{\mathrm{M}1},\, \dots,\, \mathbf{C}^{\mathfrak{l}}_{\mathrm{M}i},\, \dots,\, \mathbf{C}^{\mathfrak{l}}_{\mathrm{M}s}\right]\;,
\end{aligned} \tag{4.33a}$$

$$\left.\begin{aligned}
\mathbf{A}^{\mathfrak{l}}_{\mathrm{M}i} &= \operatorname{diag}(\mathbf{A}^{\mathfrak{l}}_{\mathrm{M}i1},\, \dots,\, \mathbf{A}^{\mathfrak{l}}_{\mathrm{M}ij},\, \dots,\, \mathbf{A}^{\mathfrak{l}}_{\mathrm{M}ir_i})\;,\\
\mathbf{B}^{\mathfrak{l}*}_{\mathrm{M}i} &= \left[\mathbf{B}^{\mathfrak{l}*}_{\mathrm{M}i1},\, \dots,\, \mathbf{B}^{\mathfrak{l}*}_{\mathrm{M}ij},\, \dots,\, \mathbf{B}^{\mathfrak{l}*}_{\mathrm{M}ir_i}\right]\;,\\
\mathbf{C}^{\mathfrak{l}}_{\mathrm{M}i} &= \left[\mathbf{C}^{\mathfrak{l}}_{\mathrm{M}i1},\, \dots,\, \mathbf{C}^{\mathfrak{l}}_{\mathrm{M}ij},\, \dots,\, \mathbf{C}^{\mathfrak{l}}_{\mathrm{M}ir_i}\right]
\end{aligned}\right\}\quad \forall\, i = 1,\, \dots,\, s\;, \tag{4.33b}$$

$$\left.\begin{aligned}\mathbf{B}^{\reflectbox{J}*}_{\mathrm{M}ij} &= \left[\mathbf{b}^{\reflectbox{J}*}_{\mathrm{M}ij1}, \dots, \mathbf{b}^{\reflectbox{J}*}_{\mathrm{M}ijk}, \dots, \mathbf{b}^{\reflectbox{J}*}_{\mathrm{M}ijq_{ij}}\right] , \\ \mathbf{C}^{\reflectbox{J}}_{\mathrm{M}ij} &= \left[\mathbf{c}^{\reflectbox{J}}_{\mathrm{M}ij1}, \dots, \mathbf{c}^{\reflectbox{J}}_{\mathrm{M}ijk}, \dots, \mathbf{c}^{\reflectbox{J}}_{\mathrm{M}ijq_{ij}}\right]\end{aligned}\right\} \quad \forall\, j = 1, \dots, r_i \,, \tag{4.33c}$$

$$\mathbf{A}^{\reflectbox{J}}_{\mathrm{M}ij} = \begin{bmatrix} \lambda_{\mathrm{M}i} & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & \lambda_{\mathrm{M}i} \end{bmatrix} \in \mathbb{C}^{q_{ij} \times q_{ij}} \quad \forall\, j = 1, \dots, r_i \,, \tag{4.33d}$$

with $\mathbf{b}^{\reflectbox{J}}_{\mathrm{M}ijk} \in \mathbb{C}^{1\times m}$ *and* $\mathbf{c}^{\reflectbox{J}}_{\mathrm{M}ijk} \in \mathbb{C}^{p\times 1}$ *corresponding to the mirrored Jordan canonical form of* $\mathbf{A}^{\reflectbox{J}}_{\mathrm{M}}$.

Then the $\mathcal{H}_2$ *inner-product of* $\mathbf{G}(s)$ *and* $\mathbf{G}_{\mathrm{M}}(s)$ *is given by*

$$\langle \mathbf{G}, \mathbf{G}_{\mathrm{M}} \rangle_{\mathcal{H}_2} = -\sum_{i=1}^{s} \sum_{j=1}^{r_i} \sum_{k=1}^{q_{ij}} \sum_{\xi=1}^{k} \mathbf{c}^{\reflectbox{J}*}_{\mathrm{M}ijk}\, \mathbf{M}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})\, \mathbf{b}^{\reflectbox{J}*}_{\mathrm{M}ij\xi} \,, \tag{4.34}$$

with the moments $\mathbf{M}^{(\mu)}(-\overline{\lambda}_{\mathrm{M}i})$ *of* $\mathbf{G}(s)$ *according to Lemma 2.24.*

Proof. The proof is contained in Appendix A. ■

Note the similar structure of (4.34) in comparison with the notation via poles and residues in [6, p. 5371]. Since [6] deals with a reduced system having simple eigenvalues, (4.34) represents a generalization for an arbitrary set of eigenvalues.

4.2.2 Parametrized Family of Reduced Transfer Functions

The following is slightly adapted from [42, p. 43ff.] (which is itself based on [5]) to better fit the purposes of this work. It represents an alternative to projective MOR: instead of projecting the FOM with $\mathbf{V}$ and $\mathbf{W}$, one can choose directly a realization of the ROM as $[\mathbf{I}_q,\, \mathbf{S}_V + \mathbf{F}\,\mathbf{R},\, \mathbf{F},\, \mathbf{C}\,\mathbf{V}]$, wherein $\mathbf{F} \in \mathbb{C}^{q\times m}$ is used as design parameter. In order to become aware of the connection between those methods, consider the ROMs obtained by each method as a *family of reduced transfer functions*, parametrized in $\mathbf{W}$ or $\mathbf{F}$:

Definition 4.18 (adapted from [42, p. 43])**.** Let $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ be the strictly proper DAE-system related to the FOM and let $\mathbf{V} \in \mathbb{C}^{n\times q}$, $\mathbf{S}_V \in \mathbb{C}^{q\times q}$ and $\mathbf{R} \in \mathbb{C}^{m\times q}$ denote the interpolation matrices according to Chapter 3, such that the generalized Sylvester equation

$$\mathbf{A}\,\mathbf{V} - \mathbf{E}\,\mathbf{V}\,\mathbf{S}_V = \mathbf{B}\,\mathbf{R} \tag{4.35}$$

is fulfilled.

Then

(i) the *family of reduced transfer functions* $\mathbf{G}_{\mathrm{F}}(s)$*, parametrized in* $\mathbf{F} \in \mathbb{C}^{q\times m}$, is defined as the set $\mathbf{G}_{\mathrm{F}}(s) = \mathbf{C}_{\mathrm{F}}\,(s\,\mathbf{E}_{\mathrm{F}} - \mathbf{A}_{\mathrm{F}})^{-1}\,\mathbf{B}_{\mathrm{F}}$, with

$$\mathbf{E}_{\mathrm{F}} = \mathbf{I}_q \,, \qquad \mathbf{A}_{\mathrm{F}} = \mathbf{S}_V + \mathbf{F}\,\mathbf{R} \,, \qquad \mathbf{B}_{\mathrm{F}} = \mathbf{F} \,, \qquad \mathbf{C}_{\mathrm{F}} = \mathbf{C}\,\mathbf{V} \,, \tag{4.36}$$

and

(ii) the *family of reduced transfer functions* $\mathbf{G}_{\mathrm{H}}(s)$*, parametrized in* $\mathbf{W} \in \mathbb{C}^{n\times q}$, is defined as the set $\mathbf{G}_{\mathrm{H}}(s) = \mathbf{C}_{\mathrm{H}}\,(s\,\mathbf{E}_{\mathrm{H}} - \mathbf{A}_{\mathrm{H}})^{-1}\,\mathbf{B}_{\mathrm{H}}$, with

$$\mathbf{E}_{\mathrm{H}} = \mathbf{W}^*\,\mathbf{E}\,\mathbf{V}\,, \qquad \mathbf{A}_{\mathrm{H}} = \mathbf{W}^*\,\mathbf{A}\,\mathbf{V}\,, \qquad \mathbf{B}_{\mathrm{H}} = \mathbf{W}^*\,\mathbf{B}\,, \qquad \mathbf{C}_{\mathrm{H}} = \mathbf{C}\,\mathbf{V}\,. \tag{4.37}$$

To analyze the relation of the two MOR methods, the generalized Sylvester equation for the interpolation data (4.35) is reformulated in Lemma 4.19.

Lemma 4.19 (adapted from [42, p. 43])**.** *Let all conditions of Definition 4.18 hold. If* $[\mathbf{E}_{\mathrm{H}},\, \mathbf{A}_{\mathrm{H}},\, \mathbf{B}_{\mathrm{H}},\, \mathbf{C}_{\mathrm{H}}]$ *is a realization of the reduced transfer function* $\mathbf{G}_{\mathrm{H}}(s)$ *obtained by projective MOR (case (ii) in Definition 4.18), then* $\mathbf{A}_{\mathrm{H}}$ *satisfies*

$$\mathbf{A}_{\mathrm{H}} = \mathbf{E}_{\mathrm{H}}\,\mathbf{S}_V + \mathbf{B}_{\mathrm{H}}\,\mathbf{R}\,. \tag{4.38}$$

Note that Lemma 4.19 is only valid in combination with projective MOR according to Section 3.5. Using this result, a connection between the two parametrization techniques can be drawn:

Theorem 4.20 (adapted from [42, p. 45])**.** *Let* $\mathbf{G}_{\mathrm{F}}(s)$ *and* $\mathbf{G}_{\mathrm{H}}(s)$ *be families of reduced transfer functions parametrized in* $\mathbf{F}$ *and* $\mathbf{W}$ *according to Definition 4.18. Then following two statements hold:*

(i) For any $\mathbf{W}$ *such that* $\mathbf{E}_{\mathrm{H}}$ *is regular, there exists a unique* $\mathbf{F}$ *such that* $\mathbf{G}_{\mathrm{H}}(s) = \mathbf{G}_{\mathrm{F}}(s)$.

(ii) If $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ *has full column rank, then for any* $\mathbf{F}$ *there exists a* $\mathbf{W}$ *such that* $\mathbf{G}_{\mathrm{F}}(s) = \mathbf{G}_{\mathrm{H}}(s)$.

Proof. Although the proof is contained in [42], it will be repeated in order to explain the dependency on $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$: to show part (i), the choice $\mathbf{F} = \mathbf{E}_{\mathrm{H}}^{-1}\,\mathbf{B}_{\mathrm{H}}$ together with $\mathbf{A}_{\mathrm{H}} = \mathbf{E}_{\mathrm{H}}\,\mathbf{S}_V + \mathbf{B}_{\mathrm{H}}\,\mathbf{R}$ from (4.38) is sufficient. For part (ii), one has additionally to show, that there exists a $\mathbf{W}$ such that

$$\mathbf{E}_{\mathrm{H}} = \mathbf{W}^*\,\mathbf{E}\,\mathbf{V} \overset{!}{=} \mathbf{I}_q = \mathbf{E}_{\mathrm{F}} \quad \text{and} \quad \mathbf{B}_{\mathrm{H}} = \mathbf{W}^*\,\mathbf{B} \overset{!}{=} \mathbf{F} = \mathbf{B}_{\mathrm{F}}\,, \tag{4.39}$$

or equivalently

$$\mathbf{W}^*\begin{bmatrix}\mathbf{E}\,\mathbf{V} & \mathbf{B}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_q & \mathbf{F}\end{bmatrix} \quad \Leftrightarrow \quad \begin{bmatrix}\mathbf{E}\,\mathbf{V} & \mathbf{B}\end{bmatrix}^*\mathbf{W} = \begin{bmatrix}\mathbf{I}_q & \mathbf{F}\end{bmatrix}^* \tag{4.40}$$

holds, which is the case, if $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]^*$ has full row rank, i. e. $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ has full column rank. ■

In Theorem 4.20 it is shown, that a parametrization of the ROM through $\mathbf{F}$ is a generalization of (one-sided) projective MOR. In the special case that $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ has full column rank, both methods are equivalent and can be reformulated into each other.

In order to prove $\mathcal{H}_2$ pseudo-optimality, the existence of a projection matrix $\mathbf{W}$ and thus the direct mapping $\mathbf{G}_{\mathrm{F}}(s) \leftrightarrow \mathbf{G}_{\mathrm{H}}(s)$ is not needed, therefore $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ may have a column rank lower than $q+m$. Nevertheless the rank of $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ will be important for the integration into the CURE-framework in Section 4.3 such that it can not be completely ignored. In the following it is assumed that $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ has full rank, while a further discussion of this issue is shifted to Appendix C.

Remark 4.21. Using $\mathbf{V} = \mathbf{V}^{\mathrm{P}}\,\mathbf{T}$ and $\mathbf{F} = \mathbf{T}^{-1}\,\mathbf{F}^{\mathrm{P}}$ together with $\mathbf{S}_V = \mathbf{T}^{-1}\,\mathbf{S}_V^{\mathrm{P}}\,\mathbf{T}$ and $\mathbf{R} = \mathbf{R}^{\mathrm{P}}\,\mathbf{T}$ shows, that one can use the primitive form of the interpolation matrices to describe $\mathbf{G}_{\mathrm{F}}(s)$:

$$\mathbf{G}_{\mathrm{F}}(s) = \mathbf{C}\,\mathbf{V}\,(s\,\mathbf{I}_q - \mathbf{S}_V - \mathbf{F}\,\mathbf{R})^{-1}\,\mathbf{F} = \mathbf{C}\,\mathbf{V}^{\mathrm{P}}\left(s\,\mathbf{I}_q - \mathbf{S}_V^{\mathrm{P}} - \mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}}\right)^{-1}\mathbf{F}^{\mathrm{P}}\,. \tag{4.41}$$

Note that (in the general case) different parameter matrices, either $\mathbf{F}$ or $\mathbf{F}^{\mathrm{P}}$, have to be used.

For the case of projective MOR conditions for tangential interpolation have been presented in Section 3.5. Since a parametrization through $\mathbf{F}$ does not belong to this class of MOR techniques, the statements from Theorem 3.19 are not applicable. Instead it can be shown, that tangential interpolation is achieved by construction, i. e. through the special choice of $[\mathbf{E}_{\mathrm{F}},\ \mathbf{A}_{\mathrm{F}},\ \mathbf{B}_{\mathrm{F}},\ \mathbf{C}_{\mathrm{F}}]$, which is stated in the following theorem (based on [42, p. 43]). Note that several modifications and generalizations have been made to the proof in order to match the notation of tangential-input rational Krylov subspaces used in this thesis.

Theorem 4.22 (based on [42, p. 43])**.** *Let* $(\mathbf{E},\ \mathbf{A},\ \mathbf{B},\ \mathbf{C},\ \mathbf{x}_0)$ *be the DAE-system describing the FOM and let* $\mathbf{G}(s)$ *denote its transfer function. Let* $\mathbf{G}_{\mathrm{F}}(s)$ *be the family of reduced transfer functions parametrized in* $\mathbf{F}$ *according to Definition 4.18 with* $\mathbf{S}_V$ *such that* $\lambda(\mathbf{S}_V) \cap \lambda(\mathbf{E},\ \mathbf{A}) = \emptyset$ *and* $\lambda(\mathbf{S}_V) \cap \lambda(\mathbf{E}_{\mathrm{F}},\ \mathbf{A}_{\mathrm{F}}) = \emptyset$*. Then* $\mathbf{G}_{\mathrm{F}}(s)$ *tangentially interpolates* $\mathbf{G}(s)$ *as encoded in* $\mathbf{S}_V$ *and* $\mathbf{R}$*, i. e.*

$$\left.\left(\frac{\mathrm{d}^{\mu}\mathbf{G}(s)}{\mathrm{d}s^{\mu}}\right)\right|_{s=s_i}\cdot\,\mathbf{r}_{ij} = \left.\left(\frac{\mathrm{d}^{\mu}\mathbf{G}_{\mathrm{F}}(s)}{\mathrm{d}s^{\mu}}\right)\right|_{s=s_i}\cdot\,\mathbf{r}_{ij} \qquad \begin{cases} \forall\, i = 1,\, \dots,\, s\,, \\ \forall\, j = 1,\, \dots,\, r_i\,, \\ \forall\, \mu = 0,\, \dots,\, q_{ij} - 1\,. \end{cases} \tag{4.42}$$

Proof. The proof is contained in Appendix A. ■

4.2.3 $\mathcal{H}_2$ Sets, Subspaces of Transfer Functions and the Hilbert Projection Theorem

In the following essential fundamentals of $\mathcal{H}_2$ sets and subspaces are introduced, which are necessary to define $\mathcal{H}_2$ pseudo-optimality. The definitions and results stated below are adapted from [34, p. 78ff.] to match the special case of the $\mathcal{H}_2$ function space[3]. At first the term *subspace* is defined:

Definition 4.23 (adapted from [34, p. 78ff.])**.** A subset $\mathcal{M}$ of $\mathcal{H}_2$ is called *subspace* of $\mathcal{H}_2$, if $\mathbf{X} + \mathbf{Y} \in \mathcal{M}$ and $\alpha\,\mathbf{X} \in \mathcal{M}$ for all $\mathbf{X},\ \mathbf{Y} \in \mathcal{M}$ and $\alpha \in \mathbb{C}$.

Note that the terms *subset* and *subspace* describe different entities. Furthermore, *orthogonality* in the context of the $\mathcal{H}_2$ function space is introduced:

Definition 4.24 (adapted from [34, p. 78ff.])**.** Two elements $\mathbf{X}$ and $\mathbf{Y}$ of $\mathcal{H}_2$ are called *orthogonal*, if $\langle \mathbf{X},\ \mathbf{Y}\rangle_{\mathcal{H}_2} = 0$ holds.

Using this, the *orthogonal complement* of a subspace can be defined:

[3] In [34] the more general case of Hilbert spaces is handled.

Lemma 4.25 (adapted from [34, p. 78ff.])**.** *Let $\mathcal{M}$ be a closed subspace of $\mathcal{H}_2$. Then the* orthogonal complement *of $\mathcal{M}$, defined by*

$$\mathcal{M}^{\perp} := \{\mathbf{Y} \in \mathcal{H}_2 \mid \forall\, \mathbf{X} \in \mathcal{M} : \langle \mathbf{X},\, \mathbf{Y} \rangle_{\mathcal{H}_2} = 0\} \ , \tag{4.43}$$

is also a closed subspace of $\mathcal{H}_2$.

As all necessary fundamentals have been presented, *subspaces of transfer functions* are defined. These are essential for the concept of $\mathcal{H}_2$ pseudo-optimality since they describe the context in which the ROM will be $\mathcal{H}_2$ optimal. For this purpose a realization of the transfer function in mirrored Jordan canonical form is used, whose structure will be exploited later on:

ODE **Definition 4.26.** Let $\mathbf{G}_{\mathrm{M}}(s) \in \mathcal{H}_2$ be the transfer function of an asymptotically stable ODE-system with m inputs and p outputs and let $[\mathbf{E}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{C}_{\mathrm{M}}^{\reflectbox{J}}]$ be a realization of $\mathbf{G}_{\mathrm{M}}(s)$ in mirrored Jordan canonical form. Then $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\right)$ denotes the subset of $\mathcal{H}_2$ with fixed $\mathbf{E}_{\mathrm{M}}^{\reflectbox{J}} = \mathbf{I}_q$, $\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}}$, $\mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}$ and arbitrary $\hat{\mathbf{C}}_{\mathrm{M}}$, i. e.

$$\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\right) := \left\{ \hat{\mathbf{G}}_{\mathrm{M}}(s) \;\middle|\; \exists\, \hat{\mathbf{C}}_{\mathrm{M}} \in \mathbb{C}^{p \times 1} : \hat{\mathbf{G}}_{\mathrm{M}}(s) = \hat{\mathbf{C}}_{\mathrm{M}}\,(s\,\mathbf{I}_q - \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}})^{-1}\,\mathbf{B}_{\mathrm{M}}^{\reflectbox{J}} \right\} \subset \mathcal{H}_2 \ , \tag{4.44}$$

with

$$\begin{aligned}
\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}} &= \operatorname{diag}(\mathbf{A}_{\mathrm{M}1}^{\reflectbox{J}},\, \dots,\, \mathbf{A}_{\mathrm{M}i}^{\reflectbox{J}},\, \dots,\, \mathbf{A}_{\mathrm{M}s}^{\reflectbox{J}}) \ , && && \text{(4.45a)}\\
\mathbf{A}_{\mathrm{M}i}^{\reflectbox{J}} &= \operatorname{diag}(\mathbf{A}_{\mathrm{M}i1}^{\reflectbox{J}},\, \dots,\, \mathbf{A}_{\mathrm{M}ij}^{\reflectbox{J}},\, \dots,\, \mathbf{A}_{\mathrm{M}ir_i}^{\reflectbox{J}}) && \forall\, i = 1,\, \dots,\, s \ , && \text{(4.45b)}\\
\mathbf{A}_{\mathrm{M}ij}^{\reflectbox{J}} &= \begin{bmatrix} \lambda_{\mathrm{M}i} & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & \lambda_{\mathrm{M}i} \end{bmatrix} \in \mathbb{C}^{q_{ij} \times q_{ij}} && \forall\, j = 1,\, \dots,\, r_i \ , && \text{(4.45c)}
\end{aligned}$$

and $\operatorname{Re}\{\lambda_{\mathrm{M}i}\} < 0\ \forall\, i = 1,\, \dots,\, s$.

As Definition 4.26 narrows the $\mathcal{H}_2$ function space down to a small subset ($\hat{\mathbf{C}}_{\mathrm{M}}$ is the only parameter of $\hat{\mathbf{G}}_{\mathrm{M}}(s)$, all other matrices are fixed), several important properties arise:

ODE **Corollary 4.27.** *All transfer functions $\hat{\mathbf{G}}_{\mathrm{M}}(s)$ contained in $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\right)$ share the same set of eigenvalues $\{\lambda_{\mathrm{M}i}\}$.*

Proof. The statement directly follows from the fact, that all elements of $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\right)$ share the same pair $(\mathbf{E}_{\mathrm{M}}^{\reflectbox{J}} = \mathbf{I}_q,\, \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}})$. ■

ODE **Lemma 4.28.** *The set $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\right) \subset \mathcal{H}_2$ is a* closed subspace *of $\mathcal{H}_2$.*

Proof. Consider $\hat{\mathbf{G}}_{\mathrm{M}1}(s),\, \hat{\mathbf{G}}_{\mathrm{M}2}(s) \in \mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\right)$ and $\alpha \in \mathbb{C}$. Since

$$\begin{aligned}
\hat{\mathbf{G}}_{\mathrm{M}1}(s) + \hat{\mathbf{G}}_{\mathrm{M}2}(s) &= \hat{\mathbf{C}}_{\mathrm{M}1}(s\,\mathbf{I}_q - \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}})^{-1}\mathbf{B}_{\mathrm{M}}^{\reflectbox{J}} + \hat{\mathbf{C}}_{\mathrm{M}2}(s\,\mathbf{I}_q - \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}})^{-1}\mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\\
&= \left(\hat{\mathbf{C}}_{\mathrm{M}1} + \hat{\mathbf{C}}_{\mathrm{M}2}\right)(s\,\mathbf{I}_q - \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}})^{-1}\mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\\
&\Rightarrow\ \hat{\mathbf{G}}_{\mathrm{M}1}(s) + \hat{\mathbf{G}}_{\mathrm{M}2}(s) \in \mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}}\right) \ ,
\end{aligned} \tag{4.46}$$

and

$$\alpha\,\hat{\mathbf{G}}_{\mathrm{M1}}(s) = \left(\alpha\,\hat{\mathbf{C}}_{\mathrm{M1}}\right)(s\,\mathbf{I}_q - \mathbf{A}_{\mathrm{M}}^{\mathfrak{l}})^{-1}\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}} \quad \Rightarrow \alpha\,\hat{\mathbf{G}}_{\mathrm{M1}}(s) \in \mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}},\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}\right)\,, \tag{4.47}$$

the conditions of Definition 4.23 are fulfilled. Therefore $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}},\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}\right)$ is a subspace of $\mathcal{H}_2$. Because $\mathbf{E}_{\mathrm{M}}^{\mathfrak{l}}$, $\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}}$, $\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}$ and $\hat{\mathbf{C}}_{\mathrm{M}}$ are either fixed or arbitrary (unbounded), the set $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}},\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}\right) \subset \mathcal{H}_2$ is closed [34, p. 79]. ■

The property of $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}},\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}\right)$ described in Lemma 4.28 is required in order to apply the *Hilbert projection theorem*, which is stated in Theorem 4.29. For the purpose of a shorter notation, the subspace $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}},\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}\right)$ will be abbreviated through $\mathcal{G}$ within the following. Accordingly its orthogonal complement will be denoted by $\mathcal{G}^{\perp}$.

Theorem 4.29 (adapted from [34, p. 80])**.** *Let $\mathcal{G}^{\perp}$ denote the orthogonal complement of $\mathcal{G}$ according to Lemma 4.25. Then every $\mathbf{G}(s) \in \mathcal{H}_2$ has a* unique *decomposition* ODE

$$\mathbf{G}(s) = \mathbf{H}_{\mathrm{opt}}(s) + \mathbf{H}_{\mathrm{opt}}^{\perp}(s) \tag{4.48}$$

where $\mathbf{H}_{\mathrm{opt}}(s)$ and $\mathbf{H}_{\mathrm{opt}}^{\perp}(s)$ are the nearest (regarding the $\mathcal{H}_2$ norm) elements to $\mathbf{G}(s)$ in $\mathcal{G}$ and $\mathcal{G}^{\perp}$ respectively, i. e.

$$\mathbf{H}_{\mathrm{opt}}(s) = \arg \min_{\mathbf{H}\in\mathcal{G}} \|\mathbf{G} - \mathbf{H}\|_{\mathcal{H}_2}\,, \tag{4.49a}$$

$$\mathbf{H}_{\mathrm{opt}}^{\perp}(s) = \arg \min_{\mathbf{H}^{\perp}\in\mathcal{G}^{\perp}} \|\mathbf{G} - \mathbf{H}^{\perp}\|_{\mathcal{H}_2}\,. \tag{4.49b}$$

Proof. Since $\mathcal{G}$ is a closed subspace of $\mathcal{H}_2$ (Lemma 4.28), one can use the proof given in [34, p. 81]. ■

A visualization of Theorem 4.29 is given in Figure 4.2. Keep in mind, that Figure 4.2 has to be interpreted as a simplifying illustration, since the $\mathcal{H}_2$ norm of a transfer function is in no way related to the Euclidean norm in Cartesian space.

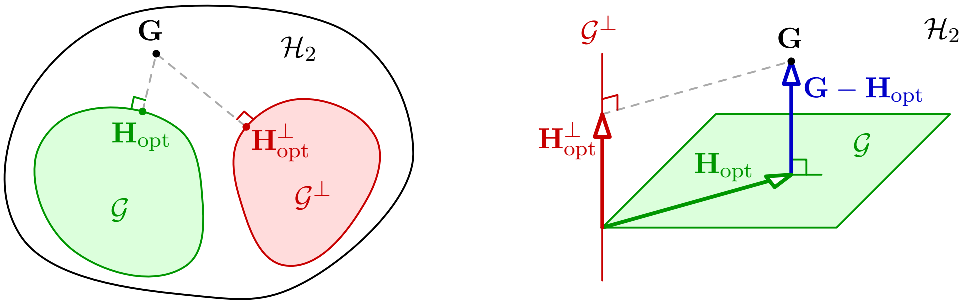

$\mathbf{G}$
$\mathcal{H}_2$
$\mathbf{H}_{\mathrm{opt}}$
$\mathcal{G}$
$\mathbf{H}_{\mathrm{opt}}^{\perp}$
$\mathcal{G}^{\perp}$
$\mathcal{G}^{\perp}$
$\mathbf{G}$
$\mathcal{H}_2$
$\mathbf{G} - \mathbf{H}_{\mathrm{opt}}$
$\mathbf{H}_{\mathrm{opt}}^{\perp}$
$\mathbf{H}_{\mathrm{opt}}$
$\mathcal{G}$

Figure 4.2: Illustration of the Hilbert projection theorem: the left side represents the optimization problems given in (4.49a) and (4.49b). An analogy for $\mathcal{H}_2 \triangleq \mathbb{R}^3$, $\mathcal{G} \triangleq \mathbb{R}^2$ and $\mathcal{G}^{\perp} \triangleq \mathbb{R}$ is depicted on the right side.

The statements of Theorem 4.29 allow to formulate an optimality condition using the $\mathcal{H}_2$ inner-product:

Corollary 4.30. *Consider the notation of Theorem 4.29. Then* ODE

$$\langle \mathbf{G} - \mathbf{H}_{\mathrm{opt}},\, \mathbf{H}\rangle_{\mathcal{H}_2} = 0 \qquad \forall\, \mathbf{H}(s) \in \mathcal{G}\,. \tag{4.50}$$

Proof. Since $\mathbf{H}_{\text{opt}}^{\perp}(s)$ is contained in $\mathcal{G}^{\perp}$, it has to be orthogonal to every $\mathbf{H}(s) \in \mathcal{G}$, i. e. $\langle \mathbf{H}_{\text{opt}}^{\perp}, \mathbf{H} \rangle_{\mathcal{H}_2} = 0 \; \forall \; \mathbf{H}(s) \in \mathcal{G}$. This leads together with $\mathbf{H}_{\text{opt}}^{\perp}(s) = \mathbf{G}(s) - \mathbf{H}_{\text{opt}}(s)$ to (4.50). ■

Note that even though Theorem 4.29 and Corollary 4.30 have been presented in the context of ODEs, only transfer functions are considered. Therefore the results are applicable to the case of *strictly proper* DAEs[4] too.

4.2.4 $\mathcal{H}_2$ Pseudo-Optimality

The basic idea of $\mathcal{H}_2$ pseudo-optimal reduction is as follows: instead of searching for the ($\mathcal{H}_2$-) optimal ROM in the set of all possible approximations (corresponding to a predefined reduced order q), one restricts oneself to a specific subset of reduced transfer functions (which was introduced as $\mathcal{G}$):

Definition 4.31 ([42, p. 80])**.** Let $\mathbf{G}(s)$ be the transfer function of the FOM and let $\mathcal{G} \subset \mathcal{H}_2$ denote a selected subset of all possible reduced transfer functions. Then the specific reduced transfer function $\mathbf{G}_{\text{r}}(s) \in \mathcal{G}$ is called $\mathcal{H}_2$ *pseudo-optimal with respect to* $\mathcal{G}$, if it satisfies

$$\mathbf{G}_{\text{r}}(s) = \arg \min_{\hat{\mathbf{G}}_{\text{M}} \in \mathcal{G}} \|\mathbf{G} - \hat{\mathbf{G}}_{\text{M}}\|_{\mathcal{H}_2} \, . \tag{4.51}$$

Note that there are infinitely many realizations, but only one *unique* transfer function $\mathbf{G}_{\text{r}}(s)$ of the $\mathcal{H}_2$ pseudo-optimal ROM.

Because of Definition 4.31 the reduced transfer function $\mathbf{G}_{\text{r}}(s)$ has to be an element of $\mathcal{G}\left(\mathbf{A}_{\text{M}}^{\mathfrak{l}}, \mathbf{B}_{\text{M}}^{\mathfrak{l}}\right)$ during $\mathcal{H}_2$ pseudo-optimal reduction. Therefore $[\mathbf{E}_{\text{M}}^{\mathfrak{l}}, \mathbf{A}_{\text{M}}^{\mathfrak{l}}, \mathbf{B}_{\text{M}}^{\mathfrak{l}}, \mathbf{C}_{\text{M}}^{\mathfrak{l}}]$ (which was used to define $\mathcal{G}\left(\mathbf{A}_{\text{M}}^{\mathfrak{l}}, \mathbf{B}_{\text{M}}^{\mathfrak{l}}\right)$ in Definition 4.26) is a valid realization of $\mathbf{G}_{\text{r}}(s)$ whose structure can be exploited to find a new formulation of Corollary 4.30:

Theorem 4.32 (extension of [42, p. 83f.])**.** *Let*

- $\mathbf{G}(s)$ *be the transfer function of an asymptotically stable and strictly proper FOM,*
- $\mathcal{G}\left(\mathbf{A}_{\text{M}}^{\mathfrak{l}}, \mathbf{B}_{\text{M}}^{\mathfrak{l}}\right)$ *be a subspace of* $\mathcal{H}_2$ *as described in Definition 4.26,*
- $\mathbf{G}_{\text{r}}(s)$ *be a specific reduced transfer function contained in* $\mathcal{G}\left(\mathbf{A}_{\text{M}}^{\mathfrak{l}}, \mathbf{B}_{\text{M}}^{\mathfrak{l}}\right)$ *and*
- $\mathbf{M}^{(\mu)}(-\overline{\lambda}_{\text{M}i})$ *and* $\mathbf{M}_{\text{r}}^{(\mu)}(-\overline{\lambda}_{\text{M}i})$ *denote the* μ*-th moments of* $\mathbf{G}(s)$ *and* $\mathbf{G}_{\text{r}}(s)$ *around* $-\overline{\lambda}_{\text{M}i}$ *respectively.*

Then $\mathbf{G}_{\text{r}}(s)$ *is the unique* $\mathcal{H}_2$ *pseudo-optimal reduced transfer function according to Definition 4.31, if and only if*

$$\langle \mathbf{G} - \mathbf{G}_{\text{r}}, \hat{\mathbf{G}}_{\text{M}} \rangle_{\mathcal{H}_2} = 0 \qquad \forall \; \hat{\mathbf{G}}_{\text{M}}(s) \in \mathcal{G}\left(\mathbf{A}_{\text{M}}^{\mathfrak{l}}, \mathbf{B}_{\text{M}}^{\mathfrak{l}}\right) \tag{4.52}$$

[4] As $\mathbf{H}(s)$ and $\mathbf{H}^{\perp}(s)$ are elements of $\mathcal{G}$ and $\mathcal{G}^{\perp}$ (thus related to ODEs), the integration of *strictly proper* DAEs is limited to realizations of $\mathbf{G}(s)$.

or equivalently

$$\sum_{\xi=1}^{k}\left(\mathbf{M}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})-\mathbf{M}_{\mathrm{r}}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})\right)\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}=\mathbf{0}\quad \textit{for all}\ \begin{cases} i=1,\,\dots,\,s\\ j=1,\,\dots,\,r_i\\ k=1,\,\dots,\,q_{ij}\end{cases}\tag{4.53}$$

holds.

Proof. Using Definition 4.31 together with (4.49a) and Corollary 4.30 shows, that $\mathbf{G}_{\mathrm{r}}(s)$ has to fulfill (4.52). Note that according to Theorem 4.29 the $\mathcal{H}_2$ pseudo-optimal reduced transfer function is unique.

To show the equivalence of (4.53), recall the result from Theorem 4.17:

$$\langle\mathbf{G}_{\mathrm{e}},\,\hat{\mathbf{G}}_{\mathrm{M}}\rangle_{\mathcal{H}_2}=-\sum_{i=1}^{s}\sum_{j=1}^{r_i}\sum_{k=1}^{q_{ij}}\sum_{\xi=1}^{k}\hat{\mathbf{c}}_{\mathrm{M}ijk}^{*}\,\mathbf{M}_{\mathrm{e}}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})\,\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}\;,\tag{4.54}$$

where $\mathbf{G}_{\mathrm{e}}(s)=\mathbf{G}(s)-\mathbf{G}_{\mathrm{r}}(s)$ denotes the error of the reduced transfer function caused by the reduction. Since

$$\begin{aligned}\mathbf{G}_{\mathrm{e}}(s)&\overset{\text{(Lemma 2.24)}}{=}-\sum_{\xi=0}^{\infty}\mathbf{M}^{(\xi)}(s_i)\,(s-s_i)^{\xi}+\sum_{\xi=0}^{\infty}\mathbf{M}_{\mathrm{r}}^{(\xi)}(s_i)\,(s-s_i)^{\xi}\\ &=-\sum_{\xi=0}^{\infty}\left(\mathbf{M}^{(\xi)}(s_i)-\mathbf{M}_{\mathrm{r}}^{(\xi)}(s_i)\right)(s-s_i)^{\xi}=-\sum_{\xi=0}^{\infty}\mathbf{M}_{\mathrm{e}}^{(\xi)}(s_i)\,(s-s_i)^{\xi}\;,\end{aligned}\tag{4.55}$$

one can find

$$\mathbf{M}_{\mathrm{e}}^{(\xi)}(s_i)=\mathbf{M}^{(\xi)}(s_i)-\mathbf{M}_{\mathrm{r}}^{(\xi)}(s_i)\;,\tag{4.56}$$

which leads to:

$$\langle\mathbf{G}-\mathbf{G}_{\mathrm{r}},\,\hat{\mathbf{G}}_{\mathrm{M}}\rangle_{\mathcal{H}_2}=-\sum_{i=1}^{s}\sum_{j=1}^{r_i}\sum_{k=1}^{q_{ij}}\sum_{\xi=1}^{k}\hat{\mathbf{c}}_{\mathrm{M}ijk}^{*}\left(\mathbf{M}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})-\mathbf{M}_{\mathrm{r}}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})\right)\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}\;.\tag{4.57}$$

Because (4.52) holds for all $\hat{\mathbf{G}}_{\mathrm{M}}(s)\in\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}},\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}\right)$ and the elements of $\mathcal{G}\left(\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}},\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}}\right)$ vary in $\hat{\mathbf{C}}_{\mathrm{M}}$, (4.57) has to be zero for arbitrary $\hat{\mathbf{c}}_{\mathrm{M}ijk}$. This leads to (4.53) and finally completes the proof[5]. ■

Through (4.53) in Theorem 4.32 a condition for $\mathcal{H}_2$ pseudo-optimality is formulated, which makes use of the moments of the involved transfer functions (FOM and ROM). Note that the moments are evaluated at the specific points $-\overline{\lambda}_{\mathrm{M}i}$ which represent the eigenvalues of $\mathbf{\Sigma}_{\mathrm{M}}$ mirrored about the imaginary axis. This will be beneficial for drawing a connection to expansion points s_i (which are chosen in the open right half of the complex plane) in the following.

As the shape of (4.53) hints, a relation to tangential interpolation is possible. For this purpose, the specific choice $[\mathbf{E}_{\mathrm{M}},\,\mathbf{A}_{\mathrm{M}},\,\mathbf{B}_{\mathrm{M}},\,\mathbf{C}_{\mathrm{M}}]=[\mathbf{I}_q,\,-\mathbf{S}_V^{*},\,\mathbf{R}^{*},\,\mathbf{C}_{\mathrm{M}}]$ is necessary ($\mathbf{C}_{\mathrm{M}}$ is still arbitrary). This leads (in addition to several beneficial properties of the ROM) to a modified condition for $\mathcal{H}_2$ pseudo-optimality, which is summarized in the following theorem:

[5] Note that the sum over ξ does not vanish, since ξ is not an index of $\hat{\mathbf{c}}_{\mathrm{M}ijk}$.

Key Theorem 4.33. *Let*

- $\mathbf{G}(s)$ *be the transfer function of an asymptotically stable and strictly proper FOM,*
- $\mathbf{S}_V^{\mathrm{P}}$*,* $\mathbf{R}^{\mathrm{P}}$ *and* $\mathbf{S}_V$*,* $\mathbf{R}$ *be the interpolation matrices related to the bases* $\mathbf{V}^{\mathrm{P}}$ *(primitive) and* $\mathbf{V}$ *(arbitrary) of* $\mathcal{K}_{\mathrm{ti}}$ *respectively, where all expansion points* s_i *are chosen in the open right half of the complex plane,*
- $[\mathbf{E}_{\mathrm{M}},\, \mathbf{A}_{\mathrm{M}},\, \mathbf{B}_{\mathrm{M}},\, \mathbf{C}_{\mathrm{M}}]$ *be a realization of the reduced transfer function* $\mathbf{G}_{\mathrm{r}}(s)$*, with the special choice*

$$\mathbf{E}_{\mathrm{M}} = \mathbf{I}_q\ , \quad \mathbf{A}_{\mathrm{M}} = -\mathbf{S}_V^{*} \quad \textit{and} \quad \mathbf{B}_{\mathrm{M}} = \mathbf{R}^{*}\ . \tag{4.58}$$

Then

(i) the ROM is asymptotically stable,

(ii) $\mathbf{G}_{\mathrm{r}}(s)$ *is contained in the subspace* $\mathcal{G}(-\mathbf{S}_V^{\mathrm{P}*},\, \mathbf{R}^{\mathrm{P}*})$*,*

(iii) the pair $(\mathbf{A}_{\mathrm{M}},\, \mathbf{B}_{\mathrm{M}})$ *is controllable and*

(iv) $\mathbf{G}_{\mathrm{r}}(s)$ *is the unique* $\mathcal{H}_2$ *pseudo-optimal reduced transfer function, if and only if it tangentially interpolates* $\mathbf{G}(s)$ *as encoded in* $\mathbf{S}_V$ *and* $\mathbf{R}$*.*

Proof. Since the expansion points lie in the open right half of the complex plane and $\mathbf{A}_{\mathrm{M}}$ is chosen such that $\mathbf{A}_{\mathrm{M}} = -\mathbf{S}_V^{*}$ with $\{s_i\} = \lambda(\mathbf{S}_V^{\mathrm{P}}) = \lambda(\mathbf{S}_V)$, all eigenvalues of the pair $(\mathbf{E}_{\mathrm{M}},\, \mathbf{A}_{\mathrm{M}})$ have negative real part, which proves part (i).

Part (ii) can be shown by inserting the transformation $\mathbf{S}_V^{*} = \mathbf{T}^{*}\,\mathbf{S}_V^{\mathrm{P}*}\,\mathbf{T}^{-*}$ and $\mathbf{R}^{*} = \mathbf{T}^{*}\,\mathbf{R}^{\mathrm{P}*}$ from Corollary 3.15 into

$$\mathbf{G}_{\mathrm{r}}(s) = \mathbf{C}_{\mathrm{M}}\,(s\,\mathbf{E}_{\mathrm{M}} - \mathbf{A}_{\mathrm{M}})^{-1}\,\mathbf{B}_{\mathrm{M}} = \mathbf{C}_{\mathrm{M}}\,(s\,\mathbf{I}_q + \mathbf{S}_V^{*})^{-1}\,\mathbf{R}^{*}\ , \tag{4.59}$$

which leads to

$$\mathbf{G}_{\mathrm{r}}(s) = \mathbf{C}_{\mathrm{M}}\,\mathbf{T}^{*}\left(s\,\mathbf{I}_q + \mathbf{S}_V^{\mathrm{P}*}\right)^{-1}\mathbf{R}^{\mathrm{P}*}\ . \tag{4.60}$$

Since $\mathbf{S}_V^{\mathrm{P}}$ is in Jordan canonical form, $-\mathbf{S}_V^{\mathrm{P}*}$ is in mirrored Jordan canonical form, thus one can find the realization $[\mathbf{E}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}},\, \mathbf{C}_{\mathrm{M}}^{\reflectbox{J}}]$ of $\mathbf{G}_{\mathrm{r}}(s)$ with

$$\mathbf{E}_{\mathrm{M}}^{\reflectbox{J}} = \mathbf{I}_q\ , \qquad \mathbf{A}_{\mathrm{M}}^{\reflectbox{J}} = -\mathbf{S}_V^{\mathrm{P}*}\ , \qquad \mathbf{B}_{\mathrm{M}}^{\reflectbox{J}} = \mathbf{R}^{\mathrm{P}*}\ , \qquad \mathbf{C}_{\mathrm{M}}^{\reflectbox{J}} = \mathbf{C}_{\mathrm{M}}\,\mathbf{T}^{*}\ , \tag{4.61}$$

which is in mirrored Jordan canonical form and defines the subspace $\mathcal{G}(-\mathbf{S}_V^{\mathrm{P}*},\, \mathbf{R}^{\mathrm{P}*})$ according to Definition 4.26.

The observability property of the pair $(\mathbf{S}_V, \mathbf{R})$ (see Section 3.4) proves part (iii):

$$(\mathbf{S}_V, \mathbf{R}) \text{ is observable } \Rightarrow \text{ rank} \begin{bmatrix} s_i\,\mathbf{I}_q - \mathbf{S}_V \\ \mathbf{R} \end{bmatrix} = q\ , \quad \forall\ s_i \in \lambda(\mathbf{S}_V)$$

$$\Rightarrow \text{ rank} \left(\underbrace{\begin{bmatrix} -\mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix}}_{\text{regular}} \begin{bmatrix} s_i\,\mathbf{I}_q - \mathbf{S}_V \\ \mathbf{R} \end{bmatrix} \right) = \text{rank} \begin{bmatrix} \mathbf{S}_V - s_i\,\mathbf{I}_q \\ \mathbf{R} \end{bmatrix} = q\ , \quad \forall\ s_i \in \lambda(\mathbf{S}_V)$$

$$\Rightarrow\ \text{rank}\begin{bmatrix}\mathbf{S}_V - s_i\,\mathbf{I}_q\\ \mathbf{R}\end{bmatrix}^* = \text{rank}\,[\mathbf{S}_V^* - \overline{s}_i\,\mathbf{I}_q\,,\ \mathbf{R}^*] = q\ , \quad \forall\ s_i \in \lambda(\mathbf{S}_V)\ ,$$

$$\Rightarrow\ \text{rank}\,[-\overline{s}_i\,\mathbf{I}_q - \mathbf{A}_\text{M}\,,\ \mathbf{B}_\text{M}] = \text{rank}\,[\lambda_{\text{M}i}\,\mathbf{I}_q - \mathbf{A}_\text{M}\,,\ \mathbf{B}_\text{M}] = q\ , \quad \forall\ \lambda_{\text{M}i} \in \lambda(\mathbf{A}_\text{M})\ ,$$

$$\Rightarrow\ (\mathbf{A}_\text{M}, \mathbf{B}_\text{M})\ \text{is controllable}\ .$$

Herein the controllability and observability criterions of Hautus [25, pp. 81,100] together with the independence of the rank regarding regular transformations (see [32, p. 9]) are used.

In order to show part (iv), the result of Theorem 4.32 is used: $\mathbf{G}_\text{r}(s)$ is the unique $\mathcal{H}_2$ pseudo-optimal reduced transfer function, if and only if

$$\sum_{\xi=1}^{k}\Big(\mathbf{M}^{(k-\xi)}(-\overline{\lambda}_{\text{M}i}) - \mathbf{M}_\text{r}^{(k-\xi)}(-\overline{\lambda}_{\text{M}i})\Big)\,\mathbf{b}_{\text{M}ij\xi}^{\mathfrak{l}*} = \mathbf{0} \quad \text{for all} \begin{cases} i = 1, \ldots, s \\ j = 1, \ldots, r_i \\ k = 1, \ldots, q_{ij} \end{cases} \tag{4.62}$$

holds. The special choice of the subspace $\mathcal{G}(-\mathbf{S}_V^{\text{P}*}, \mathbf{R}^{\text{P}*})$ leads to $-\overline{\lambda}_{\text{M}i} = s_i$ and

$$\mathbf{b}_{\text{M}ij\xi}^{\mathfrak{l}*} = \begin{cases} \mathbf{r}_{ij}\ , & \text{for } \xi = 1 \\ \mathbf{0}\ , & \text{for } \xi > 1 \end{cases}\ . \tag{4.63}$$

Thus $\mathbf{G}_\text{r}(s)$ has to fulfill

$$\mathbf{M}^{(k-1)}(s_i)\,\mathbf{r}_{ij} = \mathbf{M}_\text{r}^{(k-1)}(s_i)\,\mathbf{r}_{ij} \tag{4.64}$$

or equivalently

$$\left(\frac{\text{d}^\mu\mathbf{G}(s)}{\text{d}s^\mu}\right)\Bigg|_{s=s_i}\cdot\mathbf{r}_{ij} = \left(\frac{\text{d}^\mu\mathbf{G}_\text{r}(s)}{\text{d}s^\mu}\right)\Bigg|_{s=s_i}\cdot\mathbf{r}_{ij} \quad \text{for all} \begin{cases} i = 1, \ldots, s\ , \\ j = 1, \ldots, r_i\ , \\ \mu = 0, \ldots, q_{ij}-1 \end{cases} \tag{4.65}$$

which exactly describes tangential interpolation regarding $\mathbf{S}_V$ and $\mathbf{R}$. ■

4.2.5 Connection of $\mathbf{\Sigma}_\text{M}$, $\mathbf{\Sigma}_\text{F}$ and $\mathbf{\Sigma}_\text{r}$ to the $\mathcal{H}_2$ Pseudo-Optimal ROM

As shown in Theorem 4.33, the special choice (4.58) guarantees asymptotic stability of the ROM. Furthermore a relationship between tangential interpolation and $\mathcal{H}_2$ pseudo-optimality is formulated in part (iv) of Theorem 4.33. It is left to find the actual $\mathcal{H}_2$ pseudo-optimal ROM. For this purpose, the previous results concerning the systems $\mathbf{\Sigma}_\text{M}$, $\mathbf{\Sigma}_\text{F}$ and $\mathbf{\Sigma}_\text{r}$ (which were analyzed independently) are connected in the following.

First, the controllability and observability Gramians of the ROM related to the realization $[\mathbf{E}_\text{r},\ \mathbf{A}_\text{r},\ \mathbf{B}_\text{r},\ \mathbf{C}_\text{r}]$ are defined. Since $\mathbf{\Sigma}_\text{r}$ is assumed to be in ODE-form, the Gramians are a special case of Definition 2.32:

ODE

Definition 4.34 (derived from Definition 2.32)**.** Let $\mathbf{G}_\text{r}(s)$ be the transfer function of an asymptotically stable ROM and let $[\mathbf{E}_\text{r},\ \mathbf{A}_\text{r},\ \mathbf{B}_\text{r},\ \mathbf{C}_\text{r}]$ denote a realization of $\mathbf{G}_\text{r}(s)$ with $\det(\mathbf{E}_\text{r}) \neq 0$. Then the *controllability* and *observability Gramians* $\mathbf{\Gamma}_\text{r}^\text{c}$ and $\mathbf{\Gamma}_\text{r}^\text{o}$ related to this realization are defined as the unique Hermitian, positive semidefinite solutions of the *generalized continuous-time Lyapunov equations*

$$\mathbf{A}_\text{r}\,\mathbf{\Gamma}_\text{r}^\text{c}\,\mathbf{E}_\text{r}^* + \mathbf{E}_\text{r}\,\mathbf{\Gamma}_\text{r}^\text{c}\,\mathbf{A}_\text{r}^* + \mathbf{B}_\text{r}\,\mathbf{B}_\text{r}^* = \mathbf{0}\ , \tag{4.66}$$

and

$$\mathbf{A}_\mathrm{r}^*\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{o}\,\mathbf{E}_\mathrm{r} + \mathbf{E}_\mathrm{r}^*\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{o}\,\mathbf{A}_\mathrm{r} + \mathbf{C}_\mathrm{r}^*\,\mathbf{C}_\mathrm{r} = \mathbf{0}\;. \tag{4.67}$$

According to Definition 4.34 the Gramians are positive semidefinite, i. e. they may be singular. Since the following proofs require $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ and $\mathbf{\Gamma}_\mathrm{r}^\mathrm{o}$ to be invertible, a condition for regularity is necessary:

ODE

Lemma 4.35 (adapted from [2, pp. 72,78])**.** *Let* $[\mathbf{E}_\mathrm{r},\,\mathbf{A}_\mathrm{r},\,\mathbf{B}_\mathrm{r},\,\mathbf{C}_\mathrm{r}]$, $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ *and* $\mathbf{\Gamma}_\mathrm{r}^\mathrm{o}$ *be as in Definition 4.34. Then following statements hold:*

- *If* $(\mathbf{A}_\mathrm{r},\,\mathbf{B}_\mathrm{r})$ *is controllable, then* $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ *is positive definite, thus* $\det(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}) > 0$.
- *If* $(\mathbf{A}_\mathrm{r},\,\mathbf{C}_\mathrm{r})$ *is observable, then* $\mathbf{\Gamma}_\mathrm{r}^\mathrm{o}$ *is positive definite, thus* $\det(\mathbf{\Gamma}_\mathrm{r}^\mathrm{o}) > 0$.

Using the Gramians it can be shown, that there exists a transformation such that $[\mathbf{E}_\mathrm{M},\,\mathbf{A}_\mathrm{M},\,\mathbf{B}_\mathrm{M},\,\mathbf{C}_\mathrm{M}]$ and $[\mathbf{E}_\mathrm{r},\,\mathbf{A}_\mathrm{r},\,\mathbf{B}_\mathrm{r},\,\mathbf{C}_\mathrm{r}]$ are indeed realizations of the same reduced transfer function $\mathbf{G}_\mathrm{r}(s)$:

ODE

Lemma 4.36. *Let* $[\mathbf{E}_\mathrm{M},\,\mathbf{A}_\mathrm{M},\,\mathbf{B}_\mathrm{M},\,\mathbf{C}_\mathrm{M}]$ *with* $\mathbf{E}_\mathrm{M} = \mathbf{I}_q$ *and controllable* $(\mathbf{A}_\mathrm{M},\,\mathbf{B}_\mathrm{M})$ *be a realization of the reduced transfer function* $\mathbf{G}_\mathrm{r}(s)$*. Then* $[\mathbf{E}_\mathrm{r},\,\mathbf{A}_\mathrm{r},\,\mathbf{B}_\mathrm{r},\,\mathbf{C}_\mathrm{r}]$ *is also a realization of* $\mathbf{G}_\mathrm{r}(s)$*, if*

$$\mathbf{A}_\mathrm{r} = \mathbf{E}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{A}_\mathrm{M}\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\;, \qquad \mathbf{B}_\mathrm{r} = -\mathbf{E}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{B}_\mathrm{M}\;, \qquad \mathbf{C}_\mathrm{r} = -\mathbf{C}_\mathrm{M}\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\;, \tag{4.68}$$

whereby $\mathbf{E}_\mathrm{r} \in \mathbb{R}^{q\times q}$ *is arbitrary, such that* $\det(\mathbf{E}_\mathrm{r}) \neq 0$*, and* $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ *is the positive definite controllability Gramian according to Definition 4.34.*

Proof. Since $(\mathbf{A}_\mathrm{M},\,\mathbf{B}_\mathrm{M})$ is controllable, the controllability Gramian $\mathbf{\Gamma}_\mathrm{M}^\mathrm{c}$ of the realization $[\mathbf{E}_\mathrm{M},\,\mathbf{A}_\mathrm{M},\,\mathbf{B}_\mathrm{M},\,\mathbf{C}_\mathrm{M}]$, given through

$$\mathbf{A}_\mathrm{M}\,\mathbf{\Gamma}_\mathrm{M}^\mathrm{c}\,\mathbf{E}_\mathrm{M}^* + \mathbf{E}_\mathrm{M}\,\mathbf{\Gamma}_\mathrm{M}^\mathrm{c}\,\mathbf{A}_\mathrm{M}^* + \mathbf{B}_\mathrm{M}\,\mathbf{B}_\mathrm{M}^* = \mathbf{0}\;, \tag{4.69}$$

is positive definite (Lemma 4.35). Now assume, that $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ is regular, such that one can use (4.68) to write

$$\begin{gathered}\mathbf{A}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^* + \mathbf{E}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{A}_\mathrm{r}^* + \mathbf{B}_\mathrm{r}\,\mathbf{B}_\mathrm{r}^* = \mathbf{0}\;,\\ \overbrace{\mathbf{E}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{A}_\mathrm{M}}^{\mathbf{A}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}}\;\mathbf{E}_\mathrm{r}^* + \mathbf{E}_\mathrm{r}\;\overbrace{\mathbf{A}_\mathrm{M}^*\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*}^{\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{A}_\mathrm{r}^*} + \overbrace{\mathbf{E}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{B}_\mathrm{M}}^{-\mathbf{B}_\mathrm{r}}\;\overbrace{\mathbf{B}_\mathrm{M}^*\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*}^{-\mathbf{B}_\mathrm{r}^*} = \mathbf{0}\;,\end{gathered} \tag{4.70}$$

which leads because of $\det(\mathbf{E}_\mathrm{r}) \neq 0$ and $\mathbf{E}_\mathrm{M} = \mathbf{I}_q$ to

$$\mathbf{A}_\mathrm{M}\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{E}_\mathrm{M}^* + \mathbf{E}_\mathrm{M}\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{A}_\mathrm{M}^* + \mathbf{B}_\mathrm{M}\,\mathbf{B}_\mathrm{M}^* = \mathbf{0}\;. \tag{4.71}$$

A comparison of (4.69) with (4.71) delivers the relation $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c} = (\mathbf{\Gamma}_\mathrm{M}^\mathrm{c})^{-1}$ which verifies, that $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ is positive definite (and thus regular) too.

Finally one can easily show, that

$$\mathbf{G}_\mathrm{r}(s) = \mathbf{C}_\mathrm{M}\,(s\,\mathbf{I}_q - \mathbf{A}_\mathrm{M})^{-1}\,\mathbf{B}_\mathrm{M} = \mathbf{C}_\mathrm{r}\,(s\,\mathbf{E}_\mathrm{r} - \mathbf{A}_\mathrm{r})^{-1}\,\mathbf{B}_\mathrm{r} \tag{4.72}$$

holds for the special choice (4.68). ■

In Lemma 4.36 the transformation between the systems $\boldsymbol{\Sigma}_\mathrm{M}$ and $\boldsymbol{\Sigma}_\mathrm{r}$ was presented. If for $\boldsymbol{\Sigma}_\mathrm{M}$ the special choice given in Theorem 4.33 is made, a direct relation between $\mathbf{E}_\mathrm{r}$, $\mathbf{A}_\mathrm{r}$, $\mathbf{B}_\mathrm{r}$ and the interpolation matrices ($\mathbf{S}_V$ and $\mathbf{R}$) exists:

Corollary 4.37. *Let* $[\mathbf{E}_\mathrm{M}, \mathbf{A}_\mathrm{M}, \mathbf{B}_\mathrm{M}, \mathbf{C}_\mathrm{M}]$ *and* $[\mathbf{E}_\mathrm{r}, \mathbf{A}_\mathrm{r}, \mathbf{B}_\mathrm{r}, \mathbf{C}_\mathrm{r}]$ *be realizations of the reduced transfer function* $\mathbf{G}_\mathrm{r}(s)$ *as defined in Theorem 4.33 and Lemma 4.36. Then the equality* ODE

$$\mathbf{A}_\mathrm{r} = \mathbf{E}_\mathrm{r}\,\mathbf{S}_V + \mathbf{B}_\mathrm{r}\,\mathbf{R} \tag{4.73}$$

holds.

Proof. Using $\mathbf{A}_\mathrm{M} = -\mathbf{S}_V^*$ and $\mathbf{B}_\mathrm{M} = \mathbf{R}^*$ from Theorem 4.33 together with $\mathbf{A}_\mathrm{r} = \mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{A}_\mathrm{M}\,(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}$ and $\mathbf{B}_\mathrm{r} = -\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{B}_\mathrm{M}$ from Lemma 4.36 shows, that

$$\mathbf{A}_\mathrm{r} = -\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{S}_V^*\,(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,, \qquad \mathbf{B}_\mathrm{r} = -\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{R}^*\,. \tag{4.74}$$

holds. Inserting (4.74) into the generalized continuous-time Lyapunov equation of the controllability Gramian $\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}$ (4.66) leads together with $\boldsymbol{\Gamma}_\mathrm{r}^{\mathrm{c}*} = \boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}$ (Hermitian) to

$$\begin{aligned} \mathbf{A}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^* + \mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\overbrace{\left(-(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{S}_V\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*\right)}^{\mathbf{A}_\mathrm{r}^*} + \mathbf{B}_\mathrm{r}\,\overbrace{\left(-\mathbf{R}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*\right)}^{\mathbf{B}_\mathrm{r}^*} &= \mathbf{0} \\ (\mathbf{A}_\mathrm{r} - \mathbf{E}_\mathrm{r}\,\mathbf{S}_V - \mathbf{B}_\mathrm{r}\,\mathbf{R})\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^* &= \mathbf{0} \end{aligned} \tag{4.75}$$

which is equivalent to (4.73) because $\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}$ and $\mathbf{E}_\mathrm{r}$ are invertible. ■

Remark 4.38. Note that Corollary 4.37 is independent from the result of Lemma 4.19 since it is not restricted to projective MOR. Instead the special choice of $\mathbf{A}_\mathrm{r}$ and $\mathbf{B}_\mathrm{r}$ together with $\mathbf{A}_\mathrm{M}$ and $\mathbf{B}_\mathrm{M}$ yields the result.

Due to Theorem 4.33 the matrices $\mathbf{E}_\mathrm{M}$, $\mathbf{A}_\mathrm{M}$ and $\mathbf{B}_\mathrm{M}$ are fixed. Since $\mathbf{C}_\mathrm{M}$ has not been specified, it can be chosen such that $\mathbf{G}_\mathrm{r}(s)$ is $\mathcal{H}_2$ pseudo-optimal. The correct choice of $\mathbf{C}_\mathrm{r}$ (and thus $\mathbf{C}_\mathrm{M}$) is presented in the following theorem:

Key Theorem 4.39. *Let* $[\mathbf{E}_\mathrm{M}, \mathbf{A}_\mathrm{M}, \mathbf{B}_\mathrm{M}, \mathbf{C}_\mathrm{M}]$ *and* $[\mathbf{E}_\mathrm{r}, \mathbf{A}_\mathrm{r}, \mathbf{B}_\mathrm{r}, \mathbf{C}_\mathrm{r}]$ *be realizations of the reduced transfer function* $\mathbf{G}_\mathrm{r}(s)$ *as defined in Theorem 4.33 and Lemma 4.36. If* $\mathbf{C}_\mathrm{r}$ *is chosen such that*

$$\mathbf{C}_\mathrm{r} := \mathbf{C}\,\mathbf{V}\,, \tag{4.76}$$

then

(i) the realization $[\mathbf{E}_\mathrm{F}, \mathbf{A}_\mathrm{F}, \mathbf{B}_\mathrm{F}, \mathbf{C}_\mathrm{F}]$ *with*

$$\mathbf{E}_\mathrm{F} = \mathbf{I}_q\,, \qquad \mathbf{A}_\mathrm{F} = \mathbf{S}_V + \mathbf{F}\,\mathbf{R}\,, \qquad \mathbf{B}_\mathrm{F} = \mathbf{F}\,, \qquad \mathbf{C}_\mathrm{F} = \mathbf{C}\,\mathbf{V}\,, \tag{4.77}$$

contained in the family of reduced transfer functions $\mathbf{G}_\mathrm{F}(s)$ *parametrized in* $\mathbf{F}$ *with the special choice*

$$\mathbf{F} := \mathbf{E}_\mathrm{r}^{-1}\,\mathbf{B}_\mathrm{r}\,, \tag{4.78}$$

is a realization of $\mathbf{G}_\mathrm{r}(s)$ *too,* *(ii)* $\mathbf{G}_\mathrm{r}(s)$ *tangentially interpolates* $\mathbf{G}(s)$ *as encoded in* $\mathbf{S}_V$ *and* $\mathbf{R}$ *and*

(iii) $\mathbf{G}_\mathrm{r}(s)$ *is* $\mathcal{H}_2$ *pseudo-optimal with respect to* $\mathcal{G}(-\mathbf{S}_V^{\mathrm{P}*},\, \mathbf{R}^{\mathrm{P}*})$.

Proof. In order to show part (i), (4.78) is used:

$$\begin{aligned}\mathbf{G}_\mathrm{F}(s) &= \mathbf{C}_\mathrm{F}\,(s\,\mathbf{E}_\mathrm{F} - \mathbf{A}_\mathrm{F})^{-1}\,\mathbf{B}_\mathrm{F} = \mathbf{C}\,\mathbf{V}\left(s\,\mathbf{I}_q - \mathbf{S}_V - \mathbf{E}_\mathrm{r}^{-1}\,\mathbf{B}_\mathrm{r}\,\mathbf{R}\right)^{-1}\mathbf{E}_\mathrm{r}^{-1}\,\mathbf{B}_\mathrm{r} \\ &= \mathbf{C}\,\mathbf{V}\,(s\,\mathbf{E}_\mathrm{r} - \mathbf{E}_\mathrm{r}\,\mathbf{S}_V - \mathbf{B}_\mathrm{r}\,\mathbf{R})^{-1}\,\mathbf{B}_\mathrm{r}\ .\end{aligned} \tag{4.79}$$

Because $\mathbf{C}\,\mathbf{V} = \mathbf{C}_\mathrm{r}$ and $\mathbf{A}_\mathrm{r} = \mathbf{E}_\mathrm{r}\,\mathbf{S}_V + \mathbf{B}_\mathrm{r}\,\mathbf{R}$ (Corollary 4.37) holds, it follows

$$\mathbf{G}_\mathrm{F}(s) = \mathbf{C}_\mathrm{r}\,(s\,\mathbf{E}_\mathrm{r} - \mathbf{A}_\mathrm{r})^{-1}\,\mathbf{B}_\mathrm{r} = \mathbf{G}_\mathrm{r}(s)\ . \tag{4.80}$$

Since $\mathbf{G}_\mathrm{r}(s)$ and $\mathbf{G}_\mathrm{F}(s)$ are equal, Theorem 4.22 can be used to prove part (ii). The last part is a consequence of part (ii) and Theorem 4.33. ■

Theorem 4.39 represents the main result during the derivation of the $\mathcal{H}_2$ pseudo-optimal reduction scheme. It makes use of the three different realizations of the reduced transfer function in order to prove $\mathcal{H}_2$ pseudo-optimality, asymptotic stability and tangential interpolation of the ROM. A schematic of the relations between $\mathbf{\Sigma}_\mathrm{M}$, $\mathbf{\Sigma}_\mathrm{F}$ and $\mathbf{\Sigma}_\mathrm{r}$ is illustrated in Figure 4.3.

Remark 4.40. According to the results above, the choice of $\mathbf{E}_\mathrm{r}$ is arbitrary[6] (as long as $\det(\mathbf{E}_\mathrm{r} \neq 0)$ holds). This degree of freedom in design complies with the fact, that every transfer function has an infinite number of corresponding realizations.

4.2.6 The $\mathcal{H}_2$ Pseudo-Optimal Rational Krylov Algorithm

In the following the previous results are combined to the $\mathcal{H}_2$ pseudo-optimal rational Krylov (PORK) algorithm, which represents an efficient way to compute the $\mathcal{H}_2$ pseudo-optimal ROM corresponding to a given set of interpolation data $(\mathbf{V},\, \mathbf{S}_V,\, \mathbf{R})$. Previously additional considerations regarding the controllability Gramian $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ have to be made. Since $\mathbf{E}_\mathrm{r}$, $\mathbf{A}_\mathrm{r}$ and $\mathbf{B}_\mathrm{r}$ are not known a priori, one can not use (4.66) to compute $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ directly. Fortunately it is possible to describe $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ by the interpolation matrices $\mathbf{S}_V$ and $\mathbf{R}$:

ODE

Theorem 4.41. *Let* $[\mathbf{E}_\mathrm{M},\, \mathbf{A}_\mathrm{M},\, \mathbf{B}_\mathrm{M},\, \mathbf{C}_\mathrm{M}]$ *and* $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ *be realizations of the reduced transfer function* $\mathbf{G}_\mathrm{r}(s)$ *as defined in Theorem 4.33 and Lemma 4.36 and let* $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ *denote the controllability Gramian related to the realization* $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$*. Furthermore let all expansion points* s_i *be contained in the open right half of the complex plane.*

Then $(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}$ *is the* unique *solution of the Lyapunov equation*

$$\mathbf{S}_V^*\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1} + (\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{S}_V - \mathbf{R}^*\,\mathbf{R} = \mathbf{0}\ . \tag{4.81}$$

[6] Keep in mind, that $\mathbf{A}_\mathrm{r}$ and $\mathbf{B}_\mathrm{r}$ depend on the choice of $\mathbf{E}_\mathrm{r}$.

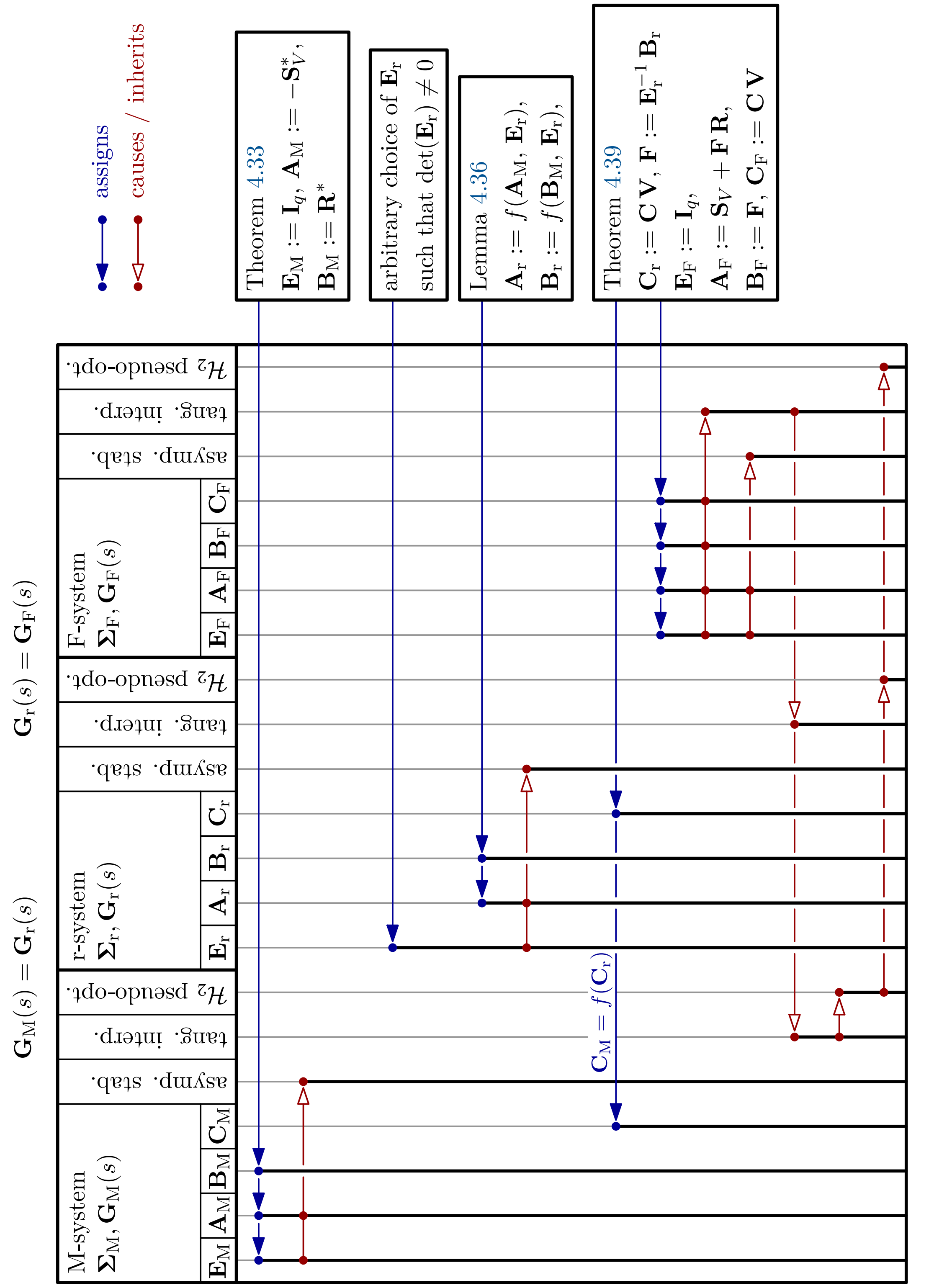

assigns
causes / inherits
Theorem 4.33
$\mathbf{E}_\mathrm{M} := \mathbf{I}_q$, $\mathbf{A}_\mathrm{M} := -\mathbf{S}_V^*$,
$\mathbf{B}_\mathrm{M} := \mathbf{R}^*$
arbitrary choice of $\mathbf{E}_\mathrm{r}$
such that $\det(\mathbf{E}_\mathrm{r}) \neq 0$
Lemma 4.36
$\mathbf{A}_\mathrm{r} := f(\mathbf{A}_\mathrm{M}, \mathbf{E}_\mathrm{r})$,
$\mathbf{B}_\mathrm{r} := f(\mathbf{B}_\mathrm{M}, \mathbf{E}_\mathrm{r})$,
Theorem 4.39
$\mathbf{C}_\mathrm{r} := \mathbf{C}\,\mathbf{V}$, $\mathbf{F} := \mathbf{E}_\mathrm{r}^{-1}\,\mathbf{B}_\mathrm{r}$
$\mathbf{E}_\mathrm{F} := \mathbf{I}_q$,
$\mathbf{A}_\mathrm{F} := \mathbf{S}_V + \mathbf{F}\,\mathbf{R}$,
$\mathbf{B}_\mathrm{F} := \mathbf{F}$, $\mathbf{C}_\mathrm{F} := \mathbf{C}\,\mathbf{V}$
$\mathbf{G}_\mathrm{M}(s) = \mathbf{G}_\mathrm{r}(s)$
$\mathbf{G}_\mathrm{r}(s) = \mathbf{G}_\mathrm{F}(s)$
M-system
$\mathbf{\Sigma}_\mathrm{M}, \mathbf{G}_\mathrm{M}(s)$
$\mathbf{E}_\mathrm{M}$
$\mathbf{A}_\mathrm{M}$
$\mathbf{B}_\mathrm{M}$
$\mathbf{C}_\mathrm{M}$
asymp. stab.
tang. interp.
$\mathcal{H}_2$ pseudo-opt.
r-system
$\mathbf{\Sigma}_\mathrm{r}, \mathbf{G}_\mathrm{r}(s)$
$\mathbf{E}_\mathrm{r}$
$\mathbf{A}_\mathrm{r}$
$\mathbf{B}_\mathrm{r}$
$\mathbf{C}_\mathrm{r}$
asymp. stab.
tang. interp.
$\mathcal{H}_2$ pseudo-opt.
F-system
$\mathbf{\Sigma}_\mathrm{F}, \mathbf{G}_\mathrm{F}(s)$
$\mathbf{E}_\mathrm{F}$
$\mathbf{A}_\mathrm{F}$
$\mathbf{B}_\mathrm{F}$
$\mathbf{C}_\mathrm{F}$
asymp. stab.
tang. interp.
$\mathcal{H}_2$ pseudo-opt.
$\mathbf{C}_\mathrm{M} = f(\mathbf{C}_\mathrm{r})$

Figure 4.3: (To be read from top to bottom in landscape format.) Dependencies between the realizations of the reduced transfer function $\mathbf{G}_\mathrm{r}(s)$ during the derivation of $\mathcal{H}_2$ pseudo-optimal reduction. A bold line indicates, that the corresponding matrix has been set or the corresponding property has been achieved.

Proof. First of all, the relations $\mathbf{A}_\mathrm{r} = -\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{S}_V^*\,(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}$ and $\mathbf{B}_\mathrm{r} = -\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{R}^*$ from Theorem 4.33 and Lemma 4.36 are inserted into (4.66) which leads to

$$\begin{aligned}\overbrace{\left(-\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{S}_V^*\,(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\right)}^{\mathbf{A}_\mathrm{r}}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^* + \mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\overbrace{\left(-(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{S}_V\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*\right)}^{\mathbf{A}_\mathrm{r}^*}& \\ + \overbrace{\left(-\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{R}^*\right)}^{\mathbf{B}_\mathrm{r}}\,\overbrace{\left(-\mathbf{R}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*\right)}^{\mathbf{B}_\mathrm{r}^*} = \mathbf{0}\,,&\end{aligned} \tag{4.82}$$

whereby $\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c} = \boldsymbol{\Gamma}_\mathrm{r}^{\mathrm{c}*}$ (Hermitian) was used. Multiplying (4.82) with $(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{E}_\mathrm{r}^{-1}$ from the left and $\mathbf{E}_\mathrm{r}^{-*}\,(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}$ from the right shows that (4.81) holds. Since $\mathrm{Re}\{s_i\} > 0$ for all $i \in \{1, \dots, s\}$, it follows that $\lambda(\mathbf{S}_V) \cap \lambda(-\mathbf{S}_V^*) = \emptyset$ and thus $(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}$ is unique according to Theorem 3.13. ■

Since all necessary tools are available, the previous results can be connected to formulate the PORK algorithm (adapted from [42, p. 91]):

Algorithm 4.1 : (input) PORK algorithm for strictly proper DAEs

```
Input : FOM: $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$ and interpolation matrices: $\mathbf{V}$, $\mathbf{S}_V$ and $\mathbf{R}$
Output : ROM: $[\mathbf{E}_\mathrm{r}, \mathbf{A}_\mathrm{r}, \mathbf{B}_\mathrm{r}, \mathbf{C}_\mathrm{r}]$
compute $(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}$ as the solution of $\mathbf{S}_V^*\,(\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1} + (\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{S}_V - \mathbf{R}^*\,\mathbf{R} = \mathbf{0}$
choose $\mathbf{E}_\mathrm{r}$ such that $\det(\mathbf{E}_\mathrm{r}) \neq 0$ (e. g. $\mathbf{E}_\mathrm{r} = \mathbf{I}_q$)
$\mathbf{B}_\mathrm{r} = -\mathbf{E}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{R}^*$        // see Theorem 4.33 and Lemma 4.36
$\mathbf{A}_\mathrm{r} = \mathbf{E}_\mathrm{r}\,\mathbf{S}_V + \mathbf{B}_\mathrm{r}\,\mathbf{R}$        // see Corollary 4.37
$\mathbf{C}_\mathrm{r} = \mathbf{C}\,\mathbf{V}$        // see Theorem 4.39
```

As Algorithm 4.1 coincides with the formulation in the ODE context given in [42, p. 91], PORK is applicable to the case of (strictly proper) DAEs without any modifications. The requirements of PORK and the properties of the resulting ROM can be summarized as follows:

Key Theorem 4.42. *Let*

- $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$ *be an asymptotically stable and strictly proper DAE-system of order n with transfer function $\mathbf{G}(s) \in \mathbb{R}^{p\times m}$ describing the FOM,*
- $\mathbf{G}(s)$ *allow realizations with real-valued system matrices,*
- $\{s_i\}$ *be a set of pairwise different expansion points contained in the open right half of the complex plane,*
- $\{\mathbf{r}_{ij}\}$ *be a set of tangential directions related to the expansion points,*
- $\mathbf{S}_V^\mathrm{P} \in \mathbb{C}^{q\times q}$ *and* $\mathbf{R}^\mathrm{P} \in \mathbb{C}^{m\times q}$ *be the (primitive) interpolation matrices containing* $\{s_i\}$ *and* $\{\mathbf{r}_{ij}\}$ *according to Section 3.2,*
- $\mathcal{K}_\mathrm{ti}$ *be the tangential-input rational Krylov subspace constructed with* $\{s_i\}$ *and* $\{\mathbf{r}_{ij}\}$ *according to Section 3.2,*
- $\mathbf{V} \in \mathbb{C}^{n\times q}$ *be an arbitrary base of* $\mathcal{K}_\mathrm{ti}$ *with corresponding interpolation matrices* $\mathbf{S}_V \in \mathbb{C}^{q\times q}$ *and* $\mathbf{R} \in \mathbb{C}^{m\times q}$ *according to Section 3.4.*

Then Algorithm 4.1 computes a realization $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ *of the ROM of order* q *in ODE-form with reduced transfer function* $\mathbf{G}_\mathrm{r}(s) \in \mathbb{R}^{p \times m}$*, such that*

(i) the system $(\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r},\, \mathbf{x}_{\mathrm{r},0})$ *is asymptotically stable,*

(ii) the eigenvalues of the ROM are the mirrored images of the expansion points, i. e. $\lambda(\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r}) = \{-\overline{s}_i\}$*,*

(iii) $\mathbf{G}_\mathrm{r}(s)$ *tangentially interpolates* $\mathbf{G}(s)$ *as encoded in* $\{s_i\}$ *and* $\{\mathbf{r}_{ij}\}$ *and*

(iv) $\mathbf{G}_\mathrm{r}(s)$ *is* $\mathcal{H}_2$ *pseudo-optimal with respect to* $\mathcal{G}\left(-\mathbf{S}_V^{\mathrm{P}*},\, \mathbf{R}^{\mathrm{P}*}\right)$.

Proof. The proofs of part (i), (iii) and (iv) are contained in Theorem 4.33 and Theorem 4.39.

In order to show part (ii) consider the computation of the eigenvalues of $(\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r})$:

$$\begin{aligned} 0 \overset{!}{=} \det(\lambda\,\mathbf{E}_\mathrm{r} - \mathbf{A}_\mathrm{r}) &= \det(\lambda\,\mathbf{E}_\mathrm{r} + \mathbf{E}_\mathrm{r}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{S}_V^*\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}) \\ &= \underbrace{\det(\mathbf{E}_\mathrm{r})}_{\neq 0}\,\underbrace{\det(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})}_{\neq 0}\,\det(\lambda\,\mathbf{I}_q + \mathbf{S}_V^*)\,\underbrace{\det((\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1})}_{\neq 0} \end{aligned} \tag{4.83}$$

and thus

$$\det(\lambda\,\mathbf{E}_\mathrm{r} - \mathbf{A}_\mathrm{r}) = 0 \quad \Leftrightarrow \quad \det(\lambda\,\mathbf{I}_q - (-\mathbf{S}_V^*)) = 0\,. \tag{4.84}$$

Since $\lambda(-\mathbf{S}_V^{\mathrm{P}*}) = \lambda(-\mathbf{S}_V^*) = \{-\overline{s}_i\}$ holds, the proof is complete. ■

Note that according to Theorem 4.42, $\lambda(\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r}) = \{-\overline{s}_i\}$ holds. This relation originates from the connection of $\mathbf{S}_V^\mathrm{P}$, which is in Jordan canonical form, and $\mathbf{A}_\mathrm{M}^\mathfrak{l}$ of the (theoretical) realization $[\mathbf{E}_\mathrm{M}^\mathfrak{l},\, \mathbf{A}_\mathrm{M}^\mathfrak{l},\, \mathbf{B}_\mathrm{M}^\mathfrak{l},\, \mathbf{C}_\mathrm{M}^\mathfrak{l}]$ of $\mathbf{G}_\mathrm{r}$, which is in *mirrored* Jordan canonical form. Part (ii) of Theorem 4.42 allows following conclusion: during $\mathcal{H}_2$ pseudo-optimal reduction with the PORK algorithm, a good choice of expansion points $\{s_i\}$ is twice as important[7], since they determine both tangential interpolation and the eigenvalues of the ROM [42, p. 103].

Although the PORK algorithm allows to directly (i. e. without iteration) compute a ROM, which is optimal in some sense, the problem of finding appropriate expansion points (and tangential directions) remains. This issue will be addressed in Section 4.3 through the SPARK algorithm.

Again, due to the duality in linear systems, all results concerning $\mathcal{H}_2$ pseudo-optimal reduction also apply to the case of tangential-output rational Krylov subspaces. For the sake of completeness the dual version of Algorithm 4.1 is stated in Algorithm 4.2 (adapted from [42, p. 92]).

Remark 4.43. Note that the matrices $\mathbf{S}_W$ and $\mathbf{L}$ differ from the notation in [42] ($\mathbf{S}_W \to \mathbf{S}_W^*$ and $\mathbf{L} \to \mathbf{L}^*$).

[7] in comparison with “usual” tangential interpolation by projective MOR

Algorithm 4.2 : (output) PORK algorithm for strictly proper DAEs

```
Input : FOM: $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ and interpolation matrices: $\mathbf{W}$, $\mathbf{S}_W$ and $\mathbf{L}$
Output : ROM: $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$
compute $(\mathbf{\Gamma}_\mathrm{r}^\mathrm{o})^{-1}$ as the solution of $\mathbf{S}_W^*\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{o})^{-1} + (\mathbf{\Gamma}_\mathrm{r}^\mathrm{o})^{-1}\,\mathbf{S}_W - \mathbf{L}^*\,\mathbf{L} = \mathbf{0}$
choose $\mathbf{E}_\mathrm{r}$ such that $\det(\mathbf{E}_\mathrm{r}) \neq 0$ (e. g. $\mathbf{E}_\mathrm{r} = \mathbf{I}_q$)
$\mathbf{C}_\mathrm{r} = -\mathbf{L}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{o}\,\mathbf{E}_r$
$\mathbf{A}_\mathrm{r} = \mathbf{S}_W^*\,\mathbf{E}_\mathrm{r} + \mathbf{L}^*\,\mathbf{C}_\mathrm{r}$
$\mathbf{B}_\mathrm{r} = \mathbf{W}^*\,\mathbf{B}$
```

4.2.7 Equivalent Conditions for $\mathcal{H}_2$ Pseudo-Optimality

In [42, p. 87ff.] seven (sufficient) conditions for $\mathcal{H}_2$ pseudo-optimality are presented. Since they are equivalent to each other, it is sufficient to enforce one of them in order to formulate a $\mathcal{H}_2$ pseudo-optimal reduction scheme (like the PORK algorithm). Keep in mind that the investigations in [42] are restricted to ODE-systems. Anyway it seems that *except for one* all conditions can be directly applied to the DAE-case. The presentation of all seven conditions would require the introduction of several additional results from [42], which is omitted for reasons of clarity and comprehensibility. Instead only the "problematic part" is discussed in the following.

According to [42, theorem 4.26 and 4.27] the equality

$$\mathbf{X} = \mathbf{V}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,, \tag{4.85}$$

with $\mathbf{X}$ as the solution of

$$\mathbf{A}\,\mathbf{X}\,\mathbf{E}_\mathrm{r}^* + \mathbf{E}\,\mathbf{X}\,\mathbf{A}_\mathrm{r}^* + \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{B}_\mathrm{r}^* = \mathbf{0}\,, \tag{4.86}$$

is equivalent to

$$\begin{gathered}\mathbf{S}_V = -\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{A}_\mathrm{r}^*\,\mathbf{E}_\mathrm{r}^{-*}\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1} \quad\Rightarrow\quad \mathbf{A}_\mathrm{r}^* = -(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{S}_V\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*\,,\\ \text{(part i) in [42, theorem 4.26])}\,,\end{gathered} \tag{4.87}$$

and

$$\begin{gathered}\mathbf{E}_\mathrm{r}^{-1}\,\mathbf{B}_\mathrm{r} + \mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{R}^* = \mathbf{0} \quad\Rightarrow\quad \mathbf{B}_\mathrm{r}^* = -\mathbf{R}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^*\,,\\ \text{(part ii) in [42, theorem 4.26])}\,,\end{gathered} \tag{4.88}$$

which are sufficient for $\mathcal{H}_2$ pseudo-optimality of $\mathbf{G}_\mathrm{r}(s)$ with respect to $\mathcal{G}\left(-\mathbf{S}_V^{\mathrm{P}*},\, \mathbf{R}^{\mathrm{P}*}\right)$. Herein the matrices $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ and $\mathbf{V}$ denote the controllability Gramian corresponding to the realization $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ of $\mathbf{G}_\mathrm{r}(s)$ and an arbitrary base of $\mathcal{K}_\mathrm{ti}$ (related to $\mathbf{S}_V$ and $\mathbf{R}$) respectively. Furthermore it is assumed, that $\mathbf{G}_\mathrm{r}(s)$ is obtained by projective MOR, i. e. $\mathbf{W}$ exists (thus $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ has full column rank). Note that in contrast to [42], the spectral projector $\mathbf{\Pi}_l^f$ of the matrix pencil $\lambda\,\mathbf{E} - \mathbf{A}$ is involved. This is because the $\mathcal{H}_2$ inner-product $\langle\mathbf{G},\, \mathbf{G}_\mathrm{r}\rangle_{\mathcal{H}_2}$ changes in the DAE-case (see Corollary 4.10).

Inserting (4.85) into (4.86) leads together with (4.87) and (4.88) to

$$\mathbf{A}\,\mathbf{V}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^* - \mathbf{E}\,\mathbf{V}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}\,\mathbf{S}_V\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^* - \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{R}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{E}_\mathrm{r}^* = \mathbf{0}\,. \tag{4.89}$$

A multiplication with $\mathbf{E}_\mathrm{r}^{-*}\,(\mathbf{\Gamma}_\mathrm{r}^\mathrm{c})^{-1}$ from the left finally results in

$$\mathbf{A}\,\mathbf{V} - \mathbf{E}\,\mathbf{V}\,\mathbf{S}_V = \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{R}\,. \tag{4.90}$$

Since $\mathbf{V}$ is defined as the *unique* solution of (3.42), the equality $\mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{R} = \mathbf{B}\,\mathbf{R}$ must hold[8]. Obviously this is fulfilled, if during the preprocessing of the improper FOM the matrix $\mathbf{B}$ is projected with $\mathbf{\Pi}_l^f$ in order to get a realization of the strictly proper subsystem, since then $\mathbf{\Pi}_l^f\mathbf{B}^\mathrm{sp} = \mathbf{B}^\mathrm{sp}$ holds. Note that this is not guaranteed in general, since the separation of the strictly proper and improper subsystems may be done by projection of $\mathbf{C}$ (or even skipped, if the FOM is strictly proper itself). This considerations are summarized in Lemma 4.44.

Lemma 4.44. *Let all conditions of Theorem 4.42 hold and* $[\mathbf{E}_\mathrm{r},\,\mathbf{A}_\mathrm{r},\,\mathbf{B}_\mathrm{r},\,\mathbf{C}_\mathrm{r}]$ *be a realization of the reduced transfer function* $\mathbf{G}_\mathrm{r}(s)$ *obtained through Algorithm 4.1. Then* $\mathbf{X}$ *as the solution of (4.86) fulfills*

$$\mathbf{X} = \mathbf{V}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}\,, \tag{4.91}$$

where $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ *denotes the controllability Gramian related to* $[\mathbf{E}_\mathrm{r},\,\mathbf{A}_\mathrm{r},\,\mathbf{B}_\mathrm{r},\,\mathbf{C}_\mathrm{r}]$*, if and only if the equality*

$$\mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{R} = \mathbf{B}\,\mathbf{R} \tag{4.92}$$

holds.

Remark 4.45. Note that the statement of Lemma 4.44 has direct influence on part ii) in [42, theorem 4.27]: if $\mathbf{X}$ and $\mathbf{V}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ are not equal, the gradient of $\|\mathbf{G}-\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2$ with respect to $\mathbf{C}_\mathrm{r}$ may not vanish (see [42, p. 148]). Nevertheless the (in)equality of $\mathbf{X}$ and $\mathbf{V}\,\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ does not affect the PORK algorithm or the properties of the resulting ROM.

4.3 Adaptive Model Order Reduction of SISO Systems

In the preceding section the application of the PORK algorithm in the context of (strictly proper) DAEs was shown. Although the ROM obtained by $\mathcal{H}_2$ pseudo-optimal reduction is optimal regarding a predefined subspace of reduced transfer functions, it is in general not $\mathcal{H}_2$ optimal. Even worse, an inappropriate choice of the interpolation data (through $\mathbf{S}_V$ and $\mathbf{R}$) may lead to a very poor approximation (despite $\mathcal{H}_2$ pseudo-optimality) [42, p. 22]. Therefore the iterative reduction scheme SPARK, which adaptively chooses suitable interpolation data for SISO-systems in combination with PORK, has been investigated in [30].

Since SPARK generates a ROM of low order, it is particularly suitable for integration into the CURE framework, which on the other hand adaptively increases the reduced order q until some kind of error tolerance is met. The reduction of (strictly proper) DAE-systems with CURE, SPARK and their combination CUREd SPARK is presented below.

Remark 4.46. Within this section SPARK and CURE are investigated for the purpose of a combination with *input* PORK (see Algorithm 4.1). Note that for all techniques dual versions for the application with *output* PORK (see Algorithm 4.2) exist. The main difference is to use (3.58) instead of (3.42).

[8] Note that although the right hand side of a generalized Sylvester equation does not influence the solvability, it contributes to the value of $\mathbf{V}$.

4.3.1 The Stability-Preserving, Adaptive Rational Krylov Algorithm

The following overview of the SPARK algorithm is summarized from [30, p. 75ff.] which is itself dedicated to the reduction of ODE-systems. Since the main algorithm remains unchanged for the case of strictly proper DAEs, detailed derivations of the results are omitted. Nevertheless several proofs (concerning the gradient and Hessian of the cost function) have to be modified in order to comply with $\det(\mathbf{E}) = 0$.

Since computing an optimal set of interpolation data is a difficult task, only SISO-systems, i. e. $m = p = 1$, are considered. As a result, the problem of finding appropriate tangential directions is avoided, thus the focus lies on the expansion points $\{s_i\}$. Furthermore the reduced order q is fixed to 2 in order to allow an efficient and robust numerical optimization (with analytical gradient and Hessian).

The main goal of SPARK is to find an optimal set $\{s_i\}$ which minimizes the error $\|\mathbf{G} - \mathbf{G}_\mathrm{r}\|^2_{\mathcal{H}_2}$ while preserving stability. This is done by setting the expansion points to

$$s_1 = a + \sqrt{a^2 - b} \quad \text{and} \quad s_2 = a - \sqrt{a^2 - b}\,, \tag{4.93}$$

where the parameters a, $b \in \mathbb{R}^{>0}$ are *adaptively* chosen during optimization. Note that since $a > 0$ holds, both expansion points lie in the open right half of the complex plane resulting in an asymptotically stable ROM according to Theorem 4.42. Moreover the parametrization via a and b allows an arbitrary choice of $\{s_i\}$ in the open right half of the complex plane:

$$\begin{aligned}
&\text{if } a^2 = b \text{ then } s_1 = s_2 \in \mathbb{R}^{>0}\,,\\
&\text{if } a^2 > b \text{ then } s_1 \neq s_2 \text{ with } s_1,\ s_2 \in \mathbb{R}^{>0}\,,\\
&\text{if } a^2 < b \text{ then } s_1 = \overline{s}_2 \in \mathbb{C} \text{ with } \mathrm{Re}\{s_1\} = \mathrm{Re}\{s_2\} > 0\,.
\end{aligned} \tag{4.94}$$

The interpolation data necessary in order to apply (input) PORK reads as

$$\begin{aligned}
&\mathbf{V} = \begin{bmatrix} \frac{1}{2}\left(\mathbf{A}_{s_1}^{-1} + \mathbf{A}_{s_2}^{-1}\right)\mathbf{B}, & \mathbf{A}_{s_2}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{B} \end{bmatrix} \in \mathbb{R}^{n\times 2}\,,\\
&\mathbf{S}_V = \begin{bmatrix} \frac{s_1+s_2}{2} & 1 \\ \left(\frac{s_1-s_2}{2}\right)^2 & \frac{s_1+s_2}{2} \end{bmatrix} = \begin{bmatrix} a & 1 \\ a^2 - b & a \end{bmatrix}\,, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 \end{bmatrix}
\end{aligned} \tag{4.95}$$

with the abbreviations $\mathbf{A}_{s_1} = (\mathbf{A} - s_1\,\mathbf{E})$ and $\mathbf{A}_{s_2} = (\mathbf{A} - s_2\,\mathbf{E})$. Note that in order to obtain a real basis $\mathbf{V}$, the transformation

$$\mathbf{T} = \begin{cases} \mathbf{I}_q & ,\ \text{if } a^2 = b\,,\\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2\sqrt{a^2-b}} \\ -\frac{1}{2} & \frac{1}{2\sqrt{a^2-b}} \end{bmatrix} & ,\ \text{if } a^2 \neq b \end{cases} \tag{4.96}$$

of the primitive basis $\mathbf{V}^\mathrm{P}$ is used (see Corollary 3.15). According to Algorithm 4.1, the realization $[\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r}]$ of the $\mathcal{H}_2$ pseudo-optimal reduced transfer function can be computed as

$$\mathbf{E}_\mathrm{r} = \mathbf{I}_2 \text{ (chosen)}\,, \quad \mathbf{A}_\mathrm{r} = \begin{bmatrix} -3\,a & 1 \\ -3\,a^2 - b & a \end{bmatrix}\,, \quad \mathbf{B}_\mathrm{r} = \begin{bmatrix} -4\,a \\ -4\,a^2 \end{bmatrix}\,, \quad \mathbf{C}_\mathrm{r} = \mathbf{C}\,\mathbf{V}\,. \tag{4.97}$$

Moreover the controllability Gramian related to $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ is given by

$$\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c} = \begin{bmatrix} 4\,a & 4\,a^2 \\ 4\,a^2 & 4\,a\,(a^2+b) \end{bmatrix} . \tag{4.98}$$

Note that the problem of minimizing the error $\|\mathbf{G}-\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2$ can be formulated in sole dependency of $\mathbf{G}_\mathrm{r}(s)$:

Lemma 4.47. *Let*

- *the FOM be described by an asymptotically stable and strictly proper DAE-system with transfer function* $\mathbf{G}(s)$*,*
- *the ROM be described by an asymptotically stable ODE-system,*
- *the reduced transfer function* $\mathbf{G}_\mathrm{r}(s)$ *be* $\mathcal{H}_2$ *pseudo-optimal with respect to a chosen subspace* $\mathcal{G}$ *and*
- $\mathbf{G}(s)$ *and* $\mathbf{G}_\mathrm{r}(s)$ *allow realizations with real-valued system matrices.*

Then minimizing the error $\|\mathbf{G}-\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2$ *is equivalent to maximizing* $\|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2$*.*

Proof. Since both transfer functions, $\mathbf{G}(s)$ and $\mathbf{G}_\mathrm{r}(s)$, allow realizations with real-valued system matrices, the equality

$$\|\mathbf{G}-\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2 \overset{(4.4)}{=} \|\mathbf{G}\|_{\mathcal{H}_2}^2 - 2\,\langle \mathbf{G},\, \mathbf{G}_\mathrm{r}\rangle_{\mathcal{H}_2} + \|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2 \tag{4.99}$$

holds. Furthermore $\mathbf{G}_\mathrm{r}(s)$ is assumed to be $\mathcal{H}_2$ pseudo-optimal, such that the inner product $\langle \mathbf{G}-\mathbf{G}_\mathrm{r},\, \mathbf{G}_\mathrm{r}\rangle_{\mathcal{H}_2}$ vanishes (see Theorem 4.32 with $\mathbf{G}_\mathrm{r} \in \mathcal{G}$). This leads to

$$0 = \langle \mathbf{G}-\mathbf{G}_\mathrm{r},\, \mathbf{G}_\mathrm{r}\rangle_{\mathcal{H}_2} = \langle \mathbf{G},\, \mathbf{G}_\mathrm{r}\rangle_{\mathcal{H}_2} - \|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2 \quad \Rightarrow \quad \langle \mathbf{G},\, \mathbf{G}_\mathrm{r}\rangle_{\mathcal{H}_2} = \|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2 \tag{4.100}$$

which is inserted into (4.99) to obtain

$$0 < \|\mathbf{G}-\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2 = \|\mathbf{G}\|_{\mathcal{H}_2}^2 - \|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2 . \tag{4.101}$$

As $\|\mathbf{G}\|_{\mathcal{H}_2}^2$ is constant (FOM), an increase of $\|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2$ implies a decrease of $\|\mathbf{G}-\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2$ which completes the proof. ■

With Lemma 4.47 and Corollary 4.11 the cost function

$$\mathcal{J} := -\|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}^2 = -\,\mathrm{tr}(\mathbf{C}_\mathrm{r}\,\boldsymbol{\Gamma}_\mathrm{r}^\mathrm{c}\,\mathbf{C}_\mathrm{r}^*) \tag{4.102}$$

can be defined and minimized by a trust-region method as proposed in [30, p. 82]. For this purpose analytic expressions of the gradient and the Hessian of $\mathcal{J}$ with respect to the parameters a and b have been derived in [30]. Note that there are few typos in [30, appendix A.1], such that the code snippet given in [30, p. 80] should be used instead.

Unfortunately the derivations of the gradient and the Hessian in [30] require the matrix $\mathbf{E}$ to be regular. Anyway all relations can be verified also for the DAE-case due to the *generalized resolvent equation* [37, p. 18]. Since the proof is rather lengthy (and requires way more effort in comparison to the ODE-case) only the most important relations are listed in Appendix D.

In Figure 4.4 the shape of the cost function for a strictly proper index 2 DAE is plotted. For demonstration, the FOM has been constructed, such that it contains (among others) a pair of complex conjugate eigenvalues at $\{\lambda_1,\, \lambda_2\} = \{-1+10\,\imath,\, -1-10\,\imath\} \in \lambda_f(\mathbf{E},\, \mathbf{A})$. This complies with the local minimum of $\mathcal{J}$ illustrated in Figure 4.4, which suggests expansion points at $\{s_1,\, s_2\} = \{1+10\,\imath,\, 1-10\,\imath\}$, i. e. at the mirrored images of $\{\lambda_1,\, \lambda_2\}$. Using this choice, the eigenvalues of the ROM $\{\lambda_{\mathrm{r},1},\, \lambda_{\mathrm{r},2}\}$ coincide with $\{\lambda_1,\, \lambda_2\}$ (since $\{\lambda_{\mathrm{r},1},\, \lambda_{\mathrm{r},2}\}$ are in turn the mirrored images of $\{s_1,\, s_2\}$ according to the PORK algorithm). Note that the local optima of $\mathcal{J}$ do not have to coincide with the mirrored images of $\lambda_f(\mathbf{E},\, \mathbf{A})$ in general.

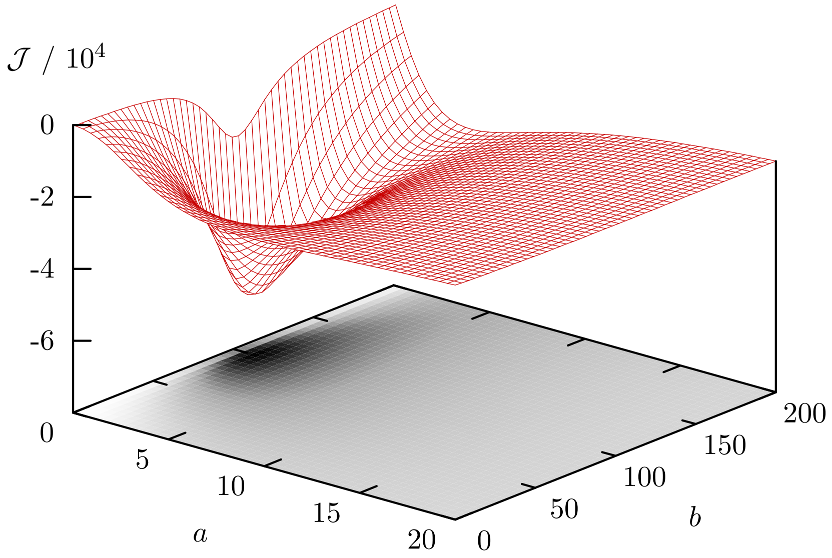


Figure 4.4: Cost function $\mathcal{J}$ in dependency of the parameters a and b for a strictly proper index 2 DAE. The (local) minimum at $a \approx 1$ and $b \approx 100$ leads to the result $s_{1,\mathrm{opt}} = 1 + 10\,\imath$ and $s_{2,\mathrm{opt}} = 1 - 10\,\imath$. Note that there may exist multiple local minima (not depicted), since the optimization problem in non-convex.

Finally Algorithm 4.3 describes SPARK as pseudo code. Note that an efficient implementation for the use with MATLAB has been presented in [30, p. 82] (therein called ESPARK).

Algorithm 4.3 : (input) SPARK algorithm for strictly proper SISO-DAEs

```
Input : FOM: [$\mathbf{E}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$]
Output : optimal interpolation matrices: $\mathbf{V}$, $\mathbf{S}_V$ and $\mathbf{R}$
choose initial values for $a,\, b \in \mathbb{R}^{>0}$
while not converged do                                        // optimization loop
    compute $\mathbf{V}$, $\mathbf{S}_V$ and $\mathbf{R}$ according to (4.95)
    compute $\mathbf{C}_\mathrm{r} = \mathbf{C}\,\mathbf{V}$ and $\mathbf{\Gamma}_\mathrm{r}^\mathrm{c}$ according to (4.98)
    evaluate the cost function $\mathcal{J}$ according to (4.102)
    compute the gradient and Hessian of $\mathcal{J}$ with respect to $(a,\, b)$ according to [30, p. 80]
    find new parameters $a,\, b \in \mathbb{R}^{>0}$, such that $\mathcal{J}$ decreases
return last choice of $\mathbf{V}$, $\mathbf{S}_V$ and $\mathbf{R}$
```

Remark 4.48. In order to reduce to computational effort of the optimization process, an advanced version of SPARK, called *Model function based Extended* SPARK (MESPARK) has been presented in [30, p. 83ff.]. The basic idea is to make use of an "intermediate" model, whose dimension is slightly greater than the one of the ROM. The *model function* is cyclically updated and replaces the FOM during the evaluation of $\mathcal{J}$ (and its gradient and Hessian) which leads to a great speed-up of the optimization process. It is worth noting that MESPARK is applicable to the case of (strictly proper) DAE-systems, if the model function is obtained through a DAE-MOR-technique. However MESPARK is not in the focus of this thesis, thus it will not be discussed further.

Note that a $\mathcal{H}_2$ pseudo-optimal ROM obtained by the PORK algorithm is in general not locally $\mathcal{H}_2$ optimal. In contrast every locally $\mathcal{H}_2$ optimal ROM is $\mathcal{H}_2$ pseudo-optimal with respect to the corresponding subspace [42, p. 96]. Using SPARK allows to circumvent the dilemma of $\mathcal{H}_2$ pseudo-optimality, i. e. the restriction to a special subspace of transfer functions. More precisely it helps to achieve (local) $\mathcal{H}_2$ optimality:

Theorem 4.49. *Let the FOM be described by an asymptotically stable and strictly proper SISO-DAE-system. Furthermore let the ROM be obtained by Algorithm 4.3 (SPARK for* $\mathbf{V}$, $\mathbf{S}_V\,\mathbf{R}$*) and Algorithm 4.1 (PORK for* $[\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r}]$*). Then the ROM is locally* $\mathcal{H}_2$ *optimal.*

Proof. Because the ROM has to be strictly proper (as the FOM is), it can be described by a realization in mirrored Jordan canonical form (see Lemma 4.16). Thus for every imaginable $\mathbf{G}_\mathrm{r}(s)$ there exists a subspace $\mathcal{G}$ (as introduced in Definition 4.26) in which it is contained. On the one hand PORK guarantees, that the ROM is (globally) $\mathcal{H}_2$ optimal within its corresponding subspace. On the other hand SPARK selects an locally optimal subspace, i. e. among all $\mathcal{H}_2$ pseudo-optimal ROMs a local minimizer of the $\mathcal{H}_2$ error is selected. ■

4.3.2 The Cumulative Reduction Framework

The following represents a brief introduction to the CURE framework and is extracted from [30, p. 57ff.] and [42, p. 49ff.]. Since all results are valid without any modifications for the case of strictly proper DAEs, only a short overview is given. For further details the interested reader is referred to [30] and [42].

The CURE framework is a technique for MOR, which assembles the ROM stepwise in a *cumulative* way. The main goal is to obtain a better approximation with each iteration, i. e. with increasing reduced order q. As the actual reduction process is not specified, it is designed to be a surrounding framework. Within the scope of this thesis CURE will be combined with SPARK (thus PORK), which is presented in the following section.

The main idea of CURE is to factorize the error system $\mathbf{G}_\mathrm{e}(s) = \mathbf{G}(s) - \mathbf{G}_\mathrm{r}(s)$. For this purpose a generalized Sylvester equation similar to (3.42) is derived:

Lemma 4.50 (adapted from [42, p. 41])**.** *Let*

- *the FOM be described by the DAE-system* $(\mathbf{E},\ \mathbf{A},\ \mathbf{B},\ \mathbf{C},\ \mathbf{x}_0)$*,*
- *the realization* $[\mathbf{E}_\mathrm{r},\ \mathbf{A}_\mathrm{r},\ \mathbf{B}_\mathrm{r},\ \mathbf{C}_\mathrm{r}]$ *of the reduced transfer function be obtained by projective MOR according to Section 3.5 such that* $\mathbf{E}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{E}\,\mathbf{V}$ *is regular and*

- *the interpolation matrices* $\mathbf{V}$, $\mathbf{S}_V$ *and* $\mathbf{R}$ *solve the generalized Sylvester equation (3.42).*

Then $\mathbf{V}$ *additionally solves the generalized Sylvester equation*

$$\mathbf{A}\,\mathbf{V} - \mathbf{E}\,\mathbf{V}\,\mathbf{E}_\mathrm{r}^{-1}\,\mathbf{A}_\mathrm{r} = \mathbf{B}_\perp\,\mathbf{R} \qquad \textit{with} \qquad \mathbf{B}_\perp := \mathbf{B} - \mathbf{E}\,\mathbf{V}\,\mathbf{E}_\mathrm{r}^{-1}\,\mathbf{B}_\mathrm{r}\ , \tag{4.103}$$

which is called $\mathbf{B}_\perp$-Sylvester equation*.*

Considering Lemma 4.50 two requirements of CURE arise. On the one hand $\mathbf{E}_\mathrm{r}$ has to be regular. This does not affect a combination with PORK, because $\det(\mathbf{E}_\mathrm{r}) \neq 0$ has to be fulfilled therein anyway (see Algorithm 4.1). On the other hand the ROM has to be obtained by projection with $\mathbf{V}$ and $\mathbf{W}$. According to Theorem 4.20 this is only equivalent to a reduction with PORK, if $[\mathbf{E}\,\mathbf{V},\,\mathbf{B}]$ has full column rank, which may not be the case in general. Note that this limitation affects both, the ODE- and DAE-case. Unfortunately it seems to be quite difficult to find universally valid conditions for $[\mathbf{E}\,\mathbf{V},\,\mathbf{B}]$ to have full column rank, such that an useful criterion has not been found yet. Nevertheless several thoughts concerning this issue (especially in the context of DAEs) are collected in Appendix C.

Using Lemma 4.50 a factorization of the error system $\mathbf{G}_\mathrm{e}(s) = \mathbf{G}(s) - \mathbf{G}_\mathrm{r}(s)$ is possible:

Theorem 4.51 ([42, p. 50])**.** *Let all conditions of Lemma 4.50 hold. Then the error model can be factorized by*

$$\mathbf{G}_\mathrm{e}(s) = \mathbf{G}(s) - \mathbf{G}_\mathrm{r}(s) = \mathbf{G}_\perp(s)\,\mathbf{G}_\mathcal{F}(s) \tag{4.104}$$

where $\mathbf{G}_\perp(s)$ *of order* n *and the feed-through model* $\mathbf{G}_\mathcal{F}(s)$ *of order* q *are defined as*

$$\mathbf{G}_\perp(s) := \mathbf{C}\,(s\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{B}_\perp\ , \tag{4.105}$$

$$\mathbf{G}_\mathcal{F}(s) := \mathbf{R}\,(s\,\mathbf{E}_\mathrm{r} - \mathbf{A}_\mathrm{r})^{-1}\,\mathbf{B}_\mathrm{r} + \mathbf{I}_m\ . \tag{4.106}$$

According to (4.104) the error $\mathbf{G}_\mathrm{e}(s)$ is equal to the product of $\mathbf{G}_\perp(s)$ and $\mathbf{G}_\mathcal{F}(s)$. Since $\mathbf{G}_\mathcal{F}(s)$ is all-pass during $\mathcal{H}_2$ pseudo-optimal reduction (see [42, p. 87]), the main dynamics of the error system is contained in $\mathbf{G}_\perp(s)$ (which is of full order n). This motivates the reduction of this factor in an additional step. After that, the approximation of $\mathbf{G}_\perp(s)$ is added (together with $\mathbf{G}_\mathcal{F}(s)$) to $\mathbf{G}_\mathrm{r}(s)$ resulting in an resized (and hopefully improved) ROM. Obviously the reduction of $\mathbf{G}_\perp(s)$ involves an error itself, which can be factorized in the same way as $\mathbf{G}_\mathrm{e}(s)$. Thus the procedure can be repeated until a desired tolerance is reached or the reduced order q hits a user-defined maximum. An illustration of the described recursion is given in Figure 4.5.

Note that the realizations of $\mathbf{G}(s)$ and $\mathbf{G}_\perp(s)$ only differ in the input matrix ($\mathbf{B} \leftrightarrow \mathbf{B}_\perp$), while $\mathbf{E}$, $\mathbf{A}$ and $\mathbf{C}$ remain unchanged. This is especially beneficial in the context of DAE-MOR, because a projection of $\mathbf{C}$ in order to separate the strictly proper and improper subsystem (see Corollary 2.23) has to be done only once. Note that a partitioning through modification of $\mathbf{B}$ seems also to be possible, since a recurring projection of $\mathbf{B}_\perp$ is (theoretically) not necessary:

Lemma 4.52. *Let all conditions of Lemma 4.50 hold. If* $\mathbf{B}$ *lives in the deflating subspace of* $\lambda\,\mathbf{E} - \mathbf{A}$ *corresponding to the finite eigenvalues, i. e.* $\mathbf{B} = \mathbf{\Pi}_l^f\,\mathbf{B}$*, so does* $\mathbf{B}_\perp$*, i. e.* $\mathbf{B}_\perp = \mathbf{\Pi}_l^f\,\mathbf{B}_\perp$*.*

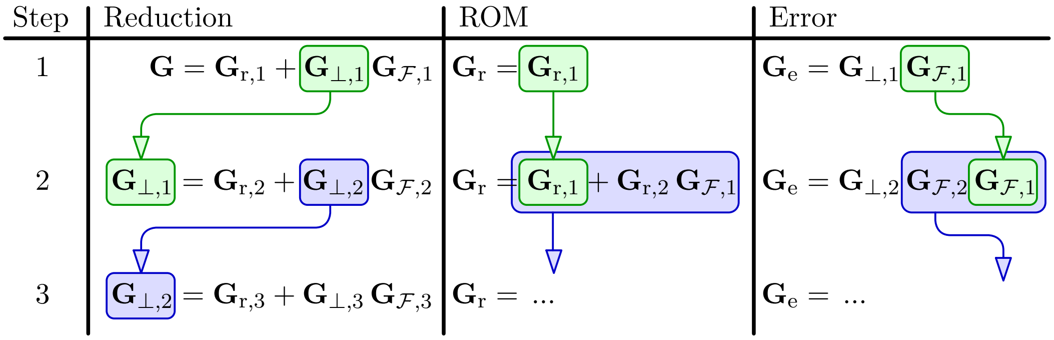

Step
Reduction
ROM
Error
1
$\mathbf{G} = \mathbf{G}_{\mathrm{r},1} + \mathbf{G}_{\perp,1}\mathbf{G}_{\mathcal{F},1}$
$\mathbf{G}_{\mathrm{r}} = \mathbf{G}_{\mathrm{r},1}$
$\mathbf{G}_{\mathrm{e}} = \mathbf{G}_{\perp,1}\mathbf{G}_{\mathcal{F},1}$
2
$\mathbf{G}_{\perp,1} = \mathbf{G}_{\mathrm{r},2} + \mathbf{G}_{\perp,2}\mathbf{G}_{\mathcal{F},2}$
$\mathbf{G}_{\mathrm{r}} = \mathbf{G}_{\mathrm{r},1} + \mathbf{G}_{\mathrm{r},2}\,\mathbf{G}_{\mathcal{F},1}$
$\mathbf{G}_{\mathrm{e}} = \mathbf{G}_{\perp,2}\mathbf{G}_{\mathcal{F},2}\mathbf{G}_{\mathcal{F},1}$
3
$\mathbf{G}_{\perp,2} = \mathbf{G}_{\mathrm{r},3} + \mathbf{G}_{\perp,3}\,\mathbf{G}_{\mathcal{F},3}$
$\mathbf{G}_{\mathrm{r}} = \ldots$
$\mathbf{G}_{\mathrm{e}} = \ldots$

Figure 4.5: Schematic of the CURE framework (inspired from [42, p. 58]). The factorization of the error is used to formulate a cumulative reduction process, such that the dimension of the reduced transfer function $\mathbf{G}_{\mathrm{r}}(s)$ grows in each iteration.

Proof. Since $\mathbf{B} = \mathbf{\Pi}_l^f\,\mathbf{B}$ holds, one can write

$$\mathbf{\Pi}_l^f\,\mathbf{B}_\perp = \mathbf{\Pi}_l^f\left(\mathbf{B} - \mathbf{E}\,\mathbf{V}\,\mathbf{E}_{\mathrm{r}}^{-1}\,\mathbf{B}_{\mathrm{r}}\right) = \underbrace{\mathbf{\Pi}_l^f\,\mathbf{B}}_{\mathbf{B}} - \mathbf{\Pi}_l^f\,\mathbf{E}\,\mathbf{V}\,\mathbf{E}_{\mathrm{r}}^{-1}\,\mathbf{B}_{\mathrm{r}}\,. \tag{4.107}$$

This leads together with $\mathbf{\Pi}_l^f\,\mathbf{E}\,\mathbf{V} = \mathbf{E}\,\mathbf{\Pi}_r^f\,\mathbf{V} = \mathbf{E}\,\mathbf{V}$ (see Lemma B.2 and Corollary B.3) to

$$\mathbf{\Pi}_l^f\,\mathbf{B}_\perp = \mathbf{B} - \mathbf{E}\,\mathbf{V}\,\mathbf{E}_{\mathrm{r}}^{-1}\,\mathbf{B}_{\mathrm{r}} = \mathbf{B}_\perp\,. \tag{4.108}$$

which completes the proof. ■

Anyway a cyclically projection of $\mathbf{B}_\perp$ might be necessary from a numerical point of view regardless of Lemma 4.52. Therefore it is assumed in the following, that the separation of the strictly proper and improper subsystems is done by modification of $\mathbf{C}$. Note that in the dual case, i. e. for interpolation with tangential-output rational Krylov subspaces, a dual version of (4.103) (involving $\mathbf{C}_\perp$) is used, such that instead a projection of $\mathbf{B}$ should be preferred.

Finally the update procedure (derived in [42, p. 54ff.]), which is executed in each iteration step of CURE, is given by:

$$\begin{aligned}
&1.)\quad \mathbf{S}_V \leftarrow \begin{bmatrix} \mathbf{S}_V & -(\mathbf{E}_{\mathrm{r}})^{-1}\,\mathbf{B}_{\mathrm{r}}\,\hat{\mathbf{R}} \\ \mathbf{0} & \hat{\mathbf{S}}_V \end{bmatrix}\,,\\
&2.)\quad \mathbf{E}_{\mathrm{r}} \leftarrow \begin{bmatrix} \mathbf{E}_{\mathrm{r}} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{E}}_{\mathrm{r}} \end{bmatrix}\,,\quad \mathbf{A}_{\mathrm{r}} \leftarrow \begin{bmatrix} \mathbf{A}_{\mathrm{r}} & \mathbf{0} \\ \hat{\mathbf{B}}_{\mathrm{r}}\,\mathbf{R} & \hat{\mathbf{A}}_{\mathrm{r}} \end{bmatrix}\,,\quad \mathbf{B}_{\mathrm{r}} \leftarrow \begin{bmatrix} \mathbf{B}_{\mathrm{r}} \\ \hat{\mathbf{B}}_{\mathrm{r}} \end{bmatrix}\,,\quad \mathbf{C}_{\mathrm{r}} \leftarrow \begin{bmatrix} \mathbf{C}_{\mathrm{r}} & \hat{\mathbf{C}}_{\mathrm{r}} \end{bmatrix}\,,\\
&3.)\quad \mathbf{V} \leftarrow \begin{bmatrix} \mathbf{V} & \hat{\mathbf{V}} \end{bmatrix}\,,\quad \mathbf{R} \leftarrow \begin{bmatrix} \mathbf{R} & \hat{\mathbf{R}} \end{bmatrix}\,,\\
&4.)\quad \mathbf{B}_\perp \leftarrow \mathbf{B}_\perp - \mathbf{E}\,\hat{\mathbf{V}}\,(\hat{\mathbf{E}}_{\mathrm{r}})^{-1}\,\hat{\mathbf{B}}_{\mathrm{r}}\,.
\end{aligned} \tag{4.109}$$

Herein $(\hat{\mathbf{V}},\,\hat{\mathbf{S}}_V,\,\hat{\mathbf{R}})$ denote the interpolation matrices used in the current reduction step, and $[\hat{\mathbf{E}}_{\mathrm{r}},\,\hat{\mathbf{A}}_{\mathrm{r}},\,\hat{\mathbf{B}}_{\mathrm{r}},\,\hat{\mathbf{C}}_{\mathrm{r}}]$ its result which is added to the ROM of the previous steps. Note that the order of the update steps in (4.109) is important (since $\mathbf{S}_V = f(\mathbf{E}_{\mathrm{r}},\,\mathbf{B}_{\mathrm{r}})$ and $\mathbf{A}_{\mathrm{r}} = f(\mathbf{R})$). Furthermore the computation of $\mathbf{V}$ and $\mathbf{S}_V$ is not necessary (since they are not used anyway), thus may be skipped. An algorithm which clarifies the overall procedure for the application with SPARK is given in the following.

4.3.3 The CUREd SPARK Algorithm

In the preceding sections SPARK and CURE have been investigated independently of each other. In Algorithm 4.4 both techniques are combined to the CUREd SPARK algorithm:

Algorithm 4.4 : (input) CUREd SPARK algorithm for strictly proper SISO-DAEs

```
Input : FOM: $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$
Output : ROM: $[\mathbf{E}_\mathrm{r}, \mathbf{A}_\mathrm{r}, \mathbf{B}_\mathrm{r}, \mathbf{C}_\mathrm{r}]$
set $\mathbf{B}_\perp = \mathbf{B}$ and $\mathbf{E}_\mathrm{r}, \mathbf{A}_\mathrm{r}, \mathbf{B}_\mathrm{r}, \mathbf{C}_\mathrm{r}, \mathbf{V}, \mathbf{S}_V, \mathbf{R} \in \mathbb{R}^{0\times 0}$    // initialization of CURE
while not converged do                                                   // iteration loop of CURE
    // find optimal interpolation matrices with SPARK
    $(\hat{\mathbf{V}}, \hat{\mathbf{S}}_V, \hat{\mathbf{R}}) = \mathrm{SPARK}(\mathbf{E}, \mathbf{A}, \mathbf{B}_\perp, \mathbf{C})$    // see Algorithm 4.3
    // use optimal interpolation matrices in PORK
    $(\hat{\mathbf{E}}_\mathrm{r}, \hat{\mathbf{A}}_\mathrm{r}, \hat{\mathbf{B}}_\mathrm{r}, \hat{\mathbf{C}}_\mathrm{r}) = \mathrm{PORK}(\mathbf{E}, \mathbf{A}, \mathbf{B}_\perp, \mathbf{C}, \hat{\mathbf{V}}, \hat{\mathbf{S}}_V, \hat{\mathbf{R}})$    // see Algorithm 4.1
    // check requirements of CURE
    if rank$[\mathbf{E}\,\hat{\mathbf{V}}, \mathbf{B}_\perp] < 2 + m$ then
        inform user
        exit while-loop
    // assemble ROM (update of $\mathbf{S}_V$ and $\mathbf{V}$ for analysis is optional)
    update $(\mathbf{S}_V)$, $\mathbf{E}_\mathrm{r}$, $\mathbf{A}_\mathrm{r}$, $\mathbf{B}_\mathrm{r}$, $\mathbf{C}_\mathrm{r}$, $(\mathbf{V})$, $\mathbf{R}$ and $\mathbf{B}_\perp$ according to (4.109)
```

Note that $\mathbf{E}_\mathrm{r}$, $\mathbf{A}_\mathrm{r}$, $\mathbf{B}_\mathrm{r}$, $\mathbf{C}_\mathrm{r}$, $\mathbf{V}$, $\mathbf{S}_V$ and $\mathbf{R}$ are initialized as empty matrices which then resize in each iteration. Furthermore $\mathbf{B}_\perp$ is set to $\mathbf{B}$ in order to force a reduction of $\mathbf{G}(s)$ in the first step. Afterwards $\mathbf{B}_\perp$ is updated cyclically such that the corresponding transfer functions $\mathbf{G}_\perp(s)$ are reduced.

Remark 4.53. In Algorithm 4.4 the column rank of $[\mathbf{E}\,\hat{\mathbf{V}}, \mathbf{B}_\perp]$ is checked. This is because in every iteration step of CURE, the reduction with PORK has to be equivalent to projective MOR with $\mathbf{W}$ and $\mathbf{V}$ (see Lemma 4.50). A corresponding $\mathbf{W}$ in turn only exists, if $[\mathbf{E}\,\hat{\mathbf{V}}, \mathbf{B}_\perp]$ has full column rank (see Theorem 4.20).

Its important to note that the ROM obtained by the CUREd SPARK algorithm is in general not locally $\mathcal{H}_2$ optimal. Although in each CURE iteration a locally $\mathcal{H}_2$ optimal reduced subsystem is obtained by SPARK and PORK (see Theorem 4.49), the overall ROM as concatenation of the subsystems does not have this property in general. Fortunately at least $\mathcal{H}_2$ pseudo-optimality is preserved as it is shown in Theorem 4.54:

Theorem 4.54 (adapted from [42, p. 101])**.** *Let the FOM be described by an asymptotically stable and strictly proper SISO-DAE-system. Furthermore let the ROM be obtained by Algorithm 4.4. Then the (overall) reduced transfer function is $\mathcal{H}_2$ pseudo-optimal and the error $\|\mathbf{G} - \mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}$ decreases monotonically with each iteration of CURE. Moreover, if $\|\hat{\mathbf{G}}_\mathrm{r}\|_{\mathcal{H}_2} \neq 0$ (related to $[\hat{\mathbf{E}}_\mathrm{r}, \hat{\mathbf{A}}_\mathrm{r}, \hat{\mathbf{B}}_\mathrm{r}, \hat{\mathbf{C}}_\mathrm{r}]$) holds for all iteration steps, then the $\mathcal{H}_2$ norm of the error decreases strictly monotonically.*

Chapter 5

Minimal Realization of the Improper Subsystem

So far methods for adaptive reduction of the strictly proper subsystem have been presented. For many technical applications in which the FOM is strictly proper by itself, this may be entirely sufficient. However, if the transfer function of the FOM involves a polynomial part $\mathbf{P}(s)$, i. e. if it is improper (or at least proper), additional effort is necessary in order to obtain good approximation results. As explained in Section 3.5 it is essential to perfectly match $\mathbf{P}(s)$, ideally by a minimal realization.

In contrast to the previous chapter, wherein reduction is done by rational Krylov subspace methods, a SVD-based approach (in particular Lyapunov BT) is used for the fast subsystem. This is because the polynomial part of the transfer function has to be matched exactly, which actually does not comply with to idea of reduction. Instead a minimal realization of $\mathbf{P}(s)$ is desired, which can be obtained by truncating non-controllable and non-observable balanced states.

First, in Section 5.1 important fundamentals are presented. After that Lyapunov BT for descriptor systems as summarized in [9] is introduced in Section 5.2. Herein an algorithm for DAE-systems (including both the strictly proper and improper subsystem) is stated. Finally the knowledge of the spectral projectors is exploited in Section 5.3 to efficiently obtain a minimal realization of the polynomial contribution $\mathbf{P}(s)$.

5.1 Fundamentals

This section represents a summary of [38] and gives a brief introduction to the fundamentals of DAE-MOR by BT. Since a detailed derivation of the following results is given in [38], only the most important relations are presented.

First of all the role of Gramians during MOR by BT is investigated. As they determine the energy behavior of the system, the Gramians are essential to obtain a balanced realization:

Lemma 5.1 (summarized from [38])**.** *Let* $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$ *be an asymptotically stable DAE-system describing the FOM whose input satisfies* $\mathbf{u}(t) = \mathbf{0}\ \forall\, t \geq 0$*. Let* $\mathbf{\Gamma}^{\text{pc}}$ *and* $\mathbf{\Gamma}^{\text{po}}$ *denote the proper controllability and observability Gramians corresponding to the realization* $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$ *as introduced in Definition 2.32 respectively. Moreover let the* *triple* $(\mathbf{E},\, \mathbf{A},\, \mathbf{B})$ *be R-controllable (see Definition 2.28) and the initial value fulfill* $\mathbf{x}_0 \in \mathcal{R}\left(\boldsymbol{\Pi}_r^f\right)$*, with* $\boldsymbol{\Pi}_r^f$ *from Definition 2.8.*

Then the future output energy can be expressed by the proper observability Gramian:

$$\int\limits_0^\infty \mathbf{y}^{\mathrm{T}}(t)\, \mathbf{y}(t)\, \mathrm{d}t = \mathbf{x}_0^{\mathrm{T}}\, \mathbf{E}^{\mathrm{T}}\, \boldsymbol{\Gamma}^{\mathrm{po}}\, \mathbf{E}\, \mathbf{x}_0 \,. \tag{5.1}$$

Furthermore the minimal past input energy that is needed to reach from $\mathbf{x}(-\infty) = \mathbf{0}$ *the state* $\mathbf{x}(0) = \mathbf{x}_0$ *is determined by the proper controllability Gramian:*

$$\min_{\mathbf{u}} \int\limits_{-\infty}^{0} \mathbf{u}^{\mathrm{T}}(t)\, \mathbf{u}(t)\, \mathrm{d}t = \mathbf{x}_0^{\mathrm{T}}\, \boldsymbol{\Gamma}^{\mathrm{pc}-}\, \mathbf{x}_0 \,, \tag{5.2}$$

wherein the matrix $\boldsymbol{\Gamma}^{\mathrm{pc}-}$ *satisfies*

$$\boldsymbol{\Gamma}^{\mathrm{pc}}\, \boldsymbol{\Gamma}^{\mathrm{pc}-}\, \boldsymbol{\Gamma}^{\mathrm{pc}} = \boldsymbol{\Gamma}^{\mathrm{pc}} \,, \qquad \boldsymbol{\Gamma}^{\mathrm{pc}-}\, \boldsymbol{\Gamma}^{\mathrm{pc}}\, \boldsymbol{\Gamma}^{\mathrm{pc}-} = \boldsymbol{\Gamma}^{\mathrm{pc}-} \,, \qquad \left(\boldsymbol{\Gamma}^{\mathrm{pc}-}\right)^* = \boldsymbol{\Gamma}^{\mathrm{pc}-} \,. \tag{5.3}$$

Note that according to Lemma 5.1 the FOM has to be asymptotically stable. This is necessary for the Gramians to exist and assumed anyway within the scope of this thesis. Furthermore (5.2) uses a pseudo-inverse of $\boldsymbol{\Gamma}^{\mathrm{pc}}$, since the actual Gramian is singular in the case of $n_\infty > 0$. This can be verified using the Weierstraß canonical form:

Lemma 5.2 (adapted from [38])**.** *Let* $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ *be an asymptotically stable DAE-system describing the FOM and let* $\mathbf{P}$ *and* $\mathbf{Q}$ *denote regular transformation matrices into Weierstraß canonical form according to Lemma 2.5. Furthermore let* $\mathbf{J}$*,* $\mathbf{N}$*,* $\tilde{\mathbf{B}}_f$*,* $\tilde{\mathbf{B}}_\infty$*,* $\tilde{\mathbf{C}}_f$ *and* $\tilde{\mathbf{C}}_\infty$ *be as stated in Definition 2.16.*

Then the proper/improper controllability/observability Gramians are partitioned as

$$\begin{aligned} \boldsymbol{\Gamma}^{\mathrm{pc}} &= \mathbf{Q} \begin{bmatrix} \boldsymbol{\Gamma}_{ff}^{\mathrm{pc}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{Q}^* \,, & \boldsymbol{\Gamma}^{\mathrm{po}} &= \mathbf{P}^* \begin{bmatrix} \boldsymbol{\Gamma}_{ff}^{\mathrm{po}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{P} \,, \\ \boldsymbol{\Gamma}^{\mathrm{imc}} &= \mathbf{Q} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imc}} \end{bmatrix} \mathbf{Q}^* \,, & \boldsymbol{\Gamma}^{\mathrm{imo}} &= \mathbf{P}^* \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imo}} \end{bmatrix} \mathbf{P} \,, \end{aligned} \tag{5.4}$$

where $\boldsymbol{\Gamma}_{ff}^{\mathrm{pc}}$*,* $\boldsymbol{\Gamma}_{ff}^{\mathrm{po}} \in \mathbb{C}^{n_f \times n_f}$ *and* $\boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imc}}$*,* $\boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imo}} \in \mathbb{C}^{n_\infty \times n_\infty}$ *satisfy the Lyapunov equations*

$$\mathbf{J}\, \boldsymbol{\Gamma}_{ff}^{\mathrm{pc}} + \boldsymbol{\Gamma}_{ff}^{\mathrm{pc}}\, \mathbf{J}^* = -\tilde{\mathbf{B}}_f\, \tilde{\mathbf{B}}_f^* \,, \tag{5.5}$$

$$\mathbf{J}^*\, \boldsymbol{\Gamma}_{ff}^{\mathrm{po}} + \boldsymbol{\Gamma}_{ff}^{\mathrm{po}}\, \mathbf{J} = -\tilde{\mathbf{C}}_f^*\, \tilde{\mathbf{C}}_f \,, \tag{5.6}$$

$$\boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imc}} - \mathbf{N}\, \boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imc}}\, \mathbf{N}^* = \tilde{\mathbf{B}}_\infty\, \tilde{\mathbf{B}}_\infty^* \,, \tag{5.7}$$

$$\boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imo}} - \mathbf{N}^*\, \boldsymbol{\Gamma}_{\infty\infty}^{\mathrm{imo}}\, \mathbf{N} = \tilde{\mathbf{C}}_\infty^*\, \tilde{\mathbf{C}}_\infty \,. \tag{5.8}$$

As (5.4) shows, the Gramians are guaranteed singular[1] in the DAE-case, i. e. for $n_f > 0$ and $n_\infty > 0$. Furthermore the similarity of (5.5) and (4.13) underlines the connection of $\mathbf{X}$ and $\mathbf{Y}$ from Lemma 4.8 with $\boldsymbol{\Gamma}^{\mathrm{pc}}$ and $\boldsymbol{\Gamma}^{\mathrm{po}}$.

Despite singularity, a “Cholesky-like” factorization is introduced:

[1] Note that this does not imply that the given realization of the FOM is not minimal.

Definition 5.3. [9, p. 9] Let $\mathbf{\Gamma}^{\mathrm{pc}}$, $\mathbf{\Gamma}^{\mathrm{po}}$, $\mathbf{\Gamma}^{\mathrm{imc}}$ and $\mathbf{\Gamma}^{\mathrm{imo}}$ denote the proper/improper controllability/observability Gramians. Then $\mathbf{\Omega}^{\mathrm{pc}}$, $\mathbf{\Omega}^{\mathrm{po}}$, $\mathbf{\Omega}^{\mathrm{imc}}$ and $\mathbf{\Omega}^{\mathrm{imo}}$ obtained by the factorizations

$$\mathbf{\Gamma}^{\mathrm{pc}} = \mathbf{\Omega}^{\mathrm{pc}}\,\mathbf{\Omega}^{\mathrm{pc}*}\,, \quad \mathbf{\Gamma}^{\mathrm{po}} = \mathbf{\Omega}^{\mathrm{po}}\,\mathbf{\Omega}^{\mathrm{po}*}\,, \quad \mathbf{\Gamma}^{\mathrm{imc}} = \mathbf{\Omega}^{\mathrm{imc}}\,\mathbf{\Omega}^{\mathrm{imc}*}\,, \quad \mathbf{\Gamma}^{\mathrm{imo}} = \mathbf{\Omega}^{\mathrm{imo}}\,\mathbf{\Omega}^{\mathrm{imo}*} \tag{5.9}$$

are called *Cholesky factors* of the corresponding Gramians.

Note that the factors in (5.9) do not represent "usual" Cholesky factors, since the requirements (in particular positive definiteness of the Gramians, see [16, p. 143]) are not satisfied. Nevertheless this naming will be used in the following.

As in the ODE-case, proper and improper Gramians are not system invariant, i. e. they are related to a corresponding realization. Because transfer functions (which represent the most important characteristic of a system during MOR) do not depend on system equivalence transformations, a direct use of the Gramians seems to be inadequate concerning the approximation of $\mathbf{G}(s)$. Instead the compounds $\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{E}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{E}$ and $\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{A}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{A}$ are considered, which introduce in some sense a system invariance that can be used:

Lemma 5.4 ([38, theorem 2.6])**.** *Let the FOM be described by the asymptotically stable DAE-system* $(\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C},\,\mathbf{x}_0)$ *and let* $\mathbf{\Gamma}^{\mathrm{pc}}$*,* $\mathbf{\Gamma}^{\mathrm{po}}$*,* $\mathbf{\Gamma}^{\mathrm{imc}}$ *and* $\mathbf{\Gamma}^{\mathrm{imo}}$ *denote the proper/improper controllability/observability Gramians related to* $[\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C}]$*. Then the matrices* $\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{E}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{E}$ *and* $\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{A}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{A}$ *are diagonalizable and have real, non-negative eigenvalues which are invariant with respect to system equivalence transformations.*

As the spectra of $\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{E}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{E}$ and $\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{A}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{A}$ are system invariant, they can be used to identify dominant contributions to the transfer function. In the context of control theory the term *Hankel singular values (HSVs)* is used:

Definition 5.5 (adapted from [38, definition 2.7])**.** Let all variables be as in Lemma 5.4. Further let n_f and n_∞ denote the dimensions of the slow and fast subsystems according to Lemma 2.5 respectively. Then

- the square roots of the largest n_f eigenvalues of the matrix $\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{E}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{E}$ denoted by θ_w^{p}, i. e.

$$\{\theta_w^{\mathrm{p}}\} = \left\{\hat{\theta}_1,\,\ldots,\,\hat{\theta}_{n_f}\;\middle|\;\hat{\theta}_w = \sqrt{\lambda_w(\mathbf{\Gamma}^{\mathrm{pc}}\,\mathbf{E}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{po}}\,\mathbf{E})}\;\wedge\;\hat{\theta}_1 \geq \hat{\theta}_2 \geq \ldots \geq 0\right\}\,, \tag{5.10}$$

  are called *proper Hankel singular values* and

- the square roots of the largest n_∞ eigenvalues of the matrix $\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{A}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{A}$ denoted by θ_w^{im}, i. e.

$$\{\theta_w^{\mathrm{im}}\} = \left\{\hat{\theta}_1,\,\ldots,\,\hat{\theta}_{n_\infty}\;\middle|\;\hat{\theta}_w = \sqrt{\lambda_w(\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{A}^{\mathrm{T}}\,\mathbf{\Gamma}^{\mathrm{imo}}\,\mathbf{A})}\;\wedge\;\hat{\theta}_1 \geq \hat{\theta}_2 \geq \ldots \geq 0\right\}\,, \tag{5.11}$$

  are called *improper Hankel singular values.*

Note that the proper HSVs represent a true generalization since they comply with the formulation of HSVs in the ODE-case (in particular for $\mathbf{E} = \mathbf{I}_n$).

The HSVs are an important measure for the contribution of the corresponding state[2] to the transfer function. Accordingly the dynamics corresponding to the comparatively small proper HSVs θ_w^{p} may be neglected without causing great error. Special care has to be taken about the improper HSVs θ_w^{im}, as they are related to the algebraic constraints of the DAE-system (which have to remain unchanged). Therefore one should truncate zero improper HSVs only, since it is guaranteed, that they do not have any influence on $\mathbf{G}(s)$.

In view of Section 5.3 an estimation for the count of non-zero improper HSVs is formulated:

Lemma 5.6 ([38])**.** *Let* $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ *be an asymptotically stable DAE-system of index* ν *describing the FOM and let* $\mathbf{\Gamma}^{\mathrm{imc}}$ *and* $\mathbf{\Gamma}^{\mathrm{imo}}$ *denote the improper controllability and observability Gramians related to the realization* $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$*. Moreover let* m *and* p *denote the count of inputs and outputs of the system respectively, while* n_∞ *represents the dimension of the fast subsystem according to Lemma 2.5.*

Then the number of non-zero improper HSVs is equal to $\operatorname{rank}(\mathbf{\Gamma}^{\mathrm{imc}}\, \mathbf{A}^{\mathrm{T}}\, \mathbf{\Gamma}^{\mathrm{imo}}\, \mathbf{A})$*, which can be estimated through*

$$\operatorname{rank}(\mathbf{\Gamma}^{\mathrm{imc}}\, \mathbf{A}^{\mathrm{T}}\, \mathbf{\Gamma}^{\mathrm{imo}}\, \mathbf{A}) \leq \min\{\nu\, m,\, \nu\, p,\, n_\infty\}\ . \tag{5.12}$$

Since in most cases $\nu\, m$ and $\nu\, p$ are small in comparison to n_∞, the order of the improper subsystem usually can be reduced significantly (without changing its contribution to the transfer function) [38].

Finally the singular value decomposition (SVD) of a matrix, which is probably the most important tool during MOR by BT, is introduced:

Lemma 5.7 (adapted from [16, pp. 70-73])**.** *Let* $\mathbf{X} \in \mathbb{C}^{u\times v}$*. Then there exist unitary matrices*

$$\mathbf{Z}_l = [\mathbf{z}_{l,1},\, \dots,\, \mathbf{z}_{l,u}] \in \mathbb{C}^{u\times u} \quad \textit{and} \quad \mathbf{Z}_r = [\mathbf{z}_{r,1},\, \dots,\, \mathbf{z}_{r,v}] \in \mathbb{C}^{v\times v} \tag{5.13}$$

such that

$$\mathbf{X} = \mathbf{Z}_l\, \mathbf{\Lambda}\, \mathbf{Z}_r^*\ , \quad \textit{with}\ \mathbf{\Lambda} = \operatorname{diag}(\lambda_1,\, \dots,\, \lambda_w) \in \mathbb{R}^{u\times v}\ , \quad w = \min\{u,\, v\}\ , \tag{5.14}$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_w \geq 0$*. The factorization* $\mathbf{X} = \mathbf{Z}_l\, \mathbf{\Lambda}\, \mathbf{Z}_r^*$ *is called* singular value decomposition.

Note that in general $\mathbf{\Lambda}$ is of rectangular shape. Since it has non-zero entries solely on its diagonal, a "trimmed" version leads to the same result:

Lemma 5.8 (adapted from [16, pp. 72])**.** *Let* $\mathbf{X} = \mathbf{Z}_l\, \mathbf{\Lambda}\, \mathbf{Z}_r^* \in \mathbb{C}^{u\times v}$ *with*

$$\begin{aligned} \mathbf{Z}_l &= [\mathbf{z}_{l,1},\, \dots,\, \mathbf{z}_{l,u}] \in \mathbb{C}^{u\times u}\ , \\ \mathbf{Z}_r &= [\mathbf{z}_{r,1},\, \dots,\, \mathbf{z}_{r,v}] \in \mathbb{C}^{v\times v}\ , \\ \mathbf{\Lambda} &= \operatorname{diag}(\lambda_1,\, \dots,\, \lambda_w) \in \mathbb{R}^{u\times v}\ , \quad w = \min\{u,\, v\} \end{aligned} \tag{5.15}$$

be the SVD of $\mathbf{X}$ *according to Lemma 5.7.*

[2] Herein the term "state" is related to the balanced realization of the FOM.

If $u > v$, then $\mathbf{X} = \hat{\mathbf{Z}}_l\,\hat{\boldsymbol{\Lambda}}\,\mathbf{Z}_r^*$ *with*

$$\begin{aligned} \hat{\mathbf{Z}}_l &= [\mathbf{z}_{l,1},\, \dots,\, \mathbf{z}_{l,v}] \in \mathbb{C}^{u\times v}\ , \\ \hat{\boldsymbol{\Lambda}} &= \operatorname{diag}(\lambda_1,\, \dots,\, \lambda_v) \in \mathbb{R}^{v\times v} \end{aligned} \tag{5.16}$$

holds. Moreover if instead $v > u$ is fulfilled, then $\mathbf{X} = \mathbf{Z}_l\,\hat{\boldsymbol{\Lambda}}\,\hat{\mathbf{Z}}_r^*$ *with*

$$\begin{aligned} \hat{\mathbf{Z}}_r &= [\mathbf{z}_{r,1},\, \dots,\, \mathbf{z}_{r,u}] \in \mathbb{C}^{v\times u}\ , \\ \hat{\boldsymbol{\Lambda}} &= \operatorname{diag}(\lambda_1,\, \dots,\, \lambda_u) \in \mathbb{R}^{u\times u} \end{aligned} \tag{5.17}$$

is satisfied. The factorizations $\mathbf{X} = \hat{\mathbf{Z}}_l\,\hat{\boldsymbol{\Lambda}}\,\mathbf{Z}_r^*$ *and* $\mathbf{X} = \mathbf{Z}_l\,\hat{\boldsymbol{\Lambda}}\,\hat{\mathbf{Z}}_r^*$ *are called* thin *singular value decompositions.*

Remark 5.9. The numerical computation of Cholesky factors as well as the SVD are general mathematical problems, which will not be discussed here. In addition, the MOR-technique presented in this thesis uses analytic expressions of the Cholesky factors $\boldsymbol{\Omega}^{\mathrm{imc}}$ and $\boldsymbol{\Omega}^{\mathrm{imo}}$ (see Section 5.3) such that dedicated algorithms are not needed.

5.2 Balanced Truncation

Since all necessary fundamentals have been presented, MOR of DAE-systems by BT is introduced in the following. This section is based on [38] and the summary given in [9, p. 8ff.] and especially treats Lyapunov BT[3] for descriptor systems.

As stated in the previous section, the (proper) Gramians describe on the one hand the output energy of the autonomous system ($\mathbf{u}(t) = \mathbf{0}\ \forall\ t \geq 0$) for a given initial value and on the other hand the minimal input energy needed to reach a specific state (see Lemma 5.1). Thus they are suited to measure the degree of controllability and observability of a specific direction in state space.

The starting point of BT is to find a *balanced* realization of $\mathbf{G}(s)$ (FOM), such that all directions in state space are as "good" controllable as observable. This is equivalent to the case, that controllability and observability Gramians are diagonal and coincide:

Definition 5.10 ([9, p. 9])**.** Let $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ be an asymptotically stable DAE-system with transfer function $\mathbf{G}(s)$ describing the FOM. Moreover let $\boldsymbol{\Gamma}^{\mathrm{pc}}$, $\boldsymbol{\Gamma}^{\mathrm{po}}$, $\boldsymbol{\Gamma}^{\mathrm{imc}}$ and $\boldsymbol{\Gamma}^{\mathrm{imo}}$ denote the proper/improper controllability/observability Gramians related to the realization $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ and $\{\theta_i^{\mathrm{p}}\}$, $\{\theta_i^{\mathrm{im}}\}$ be the corresponding sets of proper and improper HSVs according to Definition 5.5.

A realization $[\mathbf{E}_\flat,\, \mathbf{A}_\flat,\, \mathbf{B}_\flat,\, \mathbf{C}_\flat]$ of $\mathbf{G}(s)$ is called *balanced*, if the equality

$$\boldsymbol{\Gamma}_\flat^{\mathrm{pc}} + \boldsymbol{\Gamma}_\flat^{\mathrm{imc}} = \boldsymbol{\Gamma}_\flat^{\mathrm{po}} + \boldsymbol{\Gamma}_\flat^{\mathrm{imo}} = \operatorname{diag}\left(\theta_1^{\mathrm{p}},\, \dots,\, \theta_{n_f}^{\mathrm{p}},\, \theta_1^{\mathrm{im}},\, \dots,\, \theta_{n_\infty}^{\mathrm{im}}\right)\ , \tag{5.18}$$

with $\boldsymbol{\Gamma}_\flat^{\mathrm{pc}}$, $\boldsymbol{\Gamma}_\flat^{\mathrm{po}}$, $\boldsymbol{\Gamma}_\flat^{\mathrm{imc}}$ and $\boldsymbol{\Gamma}_\flat^{\mathrm{imo}}$ as the Gramians related to $[\mathbf{E}_\flat,\, \mathbf{A}_\flat,\, \mathbf{B}_\flat,\, \mathbf{C}_\flat]$, holds.

As derived in [38] the matrices $\mathbf{E}_\flat$ and $\mathbf{A}_\flat$ of a balanced realization are block diagonal:

$$\mathbf{E}_\flat = \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{\flat,\infty\infty} \end{bmatrix}\ , \qquad \mathbf{A}_\flat = \begin{bmatrix} \mathbf{A}_{\flat,ff} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}\ . \tag{5.19}$$

[3] Aside from Lyapunov BT, several variations like positive real BT (preserving stability), bounded real BT (preserving contractivity), stochastic BT and linear-quadratic Gaussian BT (for unstable systems) have been collected in [9, p. 8ff.]

Since directions, which are “poorly” controllable *and* observable, i. e. whose corresponding proper HSVs are comparatively small, do not have significant influence on the transfer function, they can be truncated. Concerning a balanced realization of $\mathbf{G}(s)$, this is equivalent to simply removing the corresponding state variables. In order to avoid the explicit computation of the balanced realization, one can make use of the SVD in order to obtain a compact implementation as given in Algorithm 5.1.

Algorithm 5.1 : Lyapunov BT for DAEs (adapted from [9, p. 10])

```
Input : FOM: $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$, spectral projectors: $\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_r^f$, $\mathbf{\Pi}_l^\infty$, $\mathbf{\Pi}_r^\infty$ and desired reduced
        order of the slow subsystem: $q_f$
Output : ROM: $[\mathbf{E}_\mathrm{r}, \mathbf{A}_\mathrm{r}, \mathbf{B}_\mathrm{r}, \mathbf{C}_\mathrm{r}]$

// balanced truncation of the slow subsystem
compute the Gramians $\mathbf{\Gamma}^\mathrm{pc}$ and $\mathbf{\Gamma}^\mathrm{po}$ satisfying (2.47) and (2.48)
compute the Cholesky factors $\mathbf{\Omega}^\mathrm{pc}$ and $\mathbf{\Omega}^\mathrm{po}$ of $\mathbf{\Gamma}^\mathrm{pc}$ and $\mathbf{\Gamma}^\mathrm{po}$ satisfying (5.9)
compute $(\mathbf{Z}_l^\mathrm{p}, \mathbf{\Lambda}^\mathrm{p}, \mathbf{Z}_r^\mathrm{p}) =$ (thin) SVD$(\mathbf{\Omega}^{\mathrm{po}*}\,\mathbf{E}\,\mathbf{\Omega}^\mathrm{pc})$          // see Lemma 5.8
truncate: $\hat{\mathbf{Z}}_l^\mathrm{p} = [\mathbf{z}_{l,1}^\mathrm{p}, \dots, \mathbf{z}_{l,q_f}^\mathrm{p}]$, $\hat{\mathbf{Z}}_r^\mathrm{p} = [\mathbf{z}_{r,1}^\mathrm{p}, \dots, \mathbf{z}_{r,q_f}^\mathrm{p}]$, $\hat{\mathbf{\Lambda}}^\mathrm{p} = \mathrm{diag}(\theta_1^\mathrm{p}, \dots, \theta_{q_f}^\mathrm{p})$

// balanced truncation of the fast subsystem
compute the Gramians $\mathbf{\Gamma}^\mathrm{imc}$ and $\mathbf{\Gamma}^\mathrm{imo}$ satisfying (2.49) and (2.50)
compute the Cholesky factors $\mathbf{\Omega}^\mathrm{imc}$ and $\mathbf{\Omega}^\mathrm{imo}$ of $\mathbf{\Gamma}^\mathrm{imc}$ and $\mathbf{\Gamma}^\mathrm{imo}$ satisfying (5.9)
compute $(\mathbf{Z}_l^\mathrm{im}, \mathbf{\Lambda}^\mathrm{im}, \mathbf{Z}_r^\mathrm{im}) =$ (thin) SVD$\big(\mathbf{\Omega}^{\mathrm{imo}*}\,\mathbf{A}\,\mathbf{\Omega}^\mathrm{imc}\big)$          // see Lemma 5.8
set $q_\infty$ as the count of non-zero improper HSVs (obtained from $\mathbf{\Lambda}^\mathrm{im}$)
truncate: $\hat{\mathbf{Z}}_l^\mathrm{im} = [\mathbf{z}_{l,1}^\mathrm{im}, \dots, \mathbf{z}_{l,q_\infty}^\mathrm{im}]$, $\hat{\mathbf{Z}}_r^\mathrm{im} = [\mathbf{z}_{r,1}^\mathrm{im}, \dots, \mathbf{z}_{r,q_\infty}^\mathrm{im}]$, $\hat{\mathbf{\Lambda}}^\mathrm{im} = \mathrm{diag}(\theta_1^\mathrm{im}, \dots, \theta_{q_\infty}^\mathrm{im})$

// assembly of the ROM by projective MOR
compute $\mathbf{W} = [\mathbf{W}_f, \mathbf{W}_\infty]$ with $\mathbf{W}_f = \mathbf{\Omega}^\mathrm{po}\,\hat{\mathbf{Z}}_l^\mathrm{p}\,(\hat{\mathbf{\Lambda}}^\mathrm{p})^{-1/2}$ and $\mathbf{W}_\infty = \mathbf{\Omega}^\mathrm{imo}\,\hat{\mathbf{Z}}_l^\mathrm{im}\,(\hat{\mathbf{\Lambda}}^\mathrm{im})^{-1/2}$
compute $\mathbf{V} = [\mathbf{V}_f, \mathbf{V}_\infty]$ with $\mathbf{V}_f = \mathbf{\Omega}^\mathrm{pc}\,\hat{\mathbf{Z}}_r^\mathrm{p}\,(\hat{\mathbf{\Lambda}}^\mathrm{p})^{-1/2}$ and $\mathbf{V}_\infty = \mathbf{\Omega}^\mathrm{imc}\,\hat{\mathbf{Z}}_r^\mathrm{im}\,(\hat{\mathbf{\Lambda}}^\mathrm{im})^{-1/2}$
project FOM: $\mathbf{E}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{E}\,\mathbf{V}$, $\mathbf{A}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{A}\,\mathbf{V}$, $\mathbf{B}_\mathrm{r} = \mathbf{W}^\mathrm{T}\,\mathbf{B}$ and $\mathbf{C}_\mathrm{r} = \mathbf{C}\,\mathbf{V}$
```

Note that the direct computation of the Cholesky factors as presented in [9, p. 22] allows to skip the formulation of the Gramians in Algorithm 5.1. For $\mathbf{\Omega}^\mathrm{imc}$ and $\mathbf{\Omega}^\mathrm{imo}$ this is done by an explicit solution formula (which will be used in Section 5.3), while $\mathbf{\Omega}^\mathrm{pc}$ and $\mathbf{\Omega}^\mathrm{po}$ have to be computed by the generalized Schur-Hammarling method, the matrix sign function method or a low rank approximation [9, p. 22f.].

Finally Theorem 5.11 summarizes the requirements of the presented method and the properties of the obtained ROM:

Key Theorem 5.11 (summarized from [9, p. 8ff.] and [38])**.** *Let*

- *the FOM be described by the asymptotically stable DAE-system* $(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$ *with transfer function* $\mathbf{G}(s)$*,*
- $\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_r^f$, $\mathbf{\Pi}_l^\infty$ *and* $\mathbf{\Pi}_r^\infty$ *denote the spectral projectors according to Definition 2.8 which are known in advance,*
- $\mathbf{G}(s)$ *be composed of* $\mathbf{G}^\mathrm{sp}(s)$ *and* $\mathbf{P}(s)$ *corresponding to the strictly proper (slow) and improper (fast) subsystem respectively, i. e.* $\mathbf{G}(s) = \mathbf{G}^\mathrm{sp}(s) + \mathbf{P}(s)$ *and*

- *q_f be the reduced order of the strictly proper subsystem which has to be chosen less than or equal to the count of non-zero proper HSVs.*

If the realization $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ of the ROM is obtained through Algorithm 5.1, then

- *$[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ is balanced,*
- *the ROM is asymptotically stable,*
- *the polynomial part of the transfer function is perfectly matched, i. e. $\mathbf{P}_\mathrm{r}(s) = \mathbf{P}(s)$,*
- *the index of the ROM $\nu_\mathrm{r} = \mathrm{ind}(\lambda\,\mathbf{E}_\mathrm{r} - \mathbf{A}_\mathrm{r})$ is equal to $\mathcal{O}(\mathbf{P}(s)) + 1$ and does not exceed the index of the FOM $\nu = \mathrm{ind}(\lambda\,\mathbf{E} - \mathbf{A})$,*
- *$[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ is a minimal realization of $\mathbf{G}_\mathrm{r}(s) = \mathbf{G}_\mathrm{r}^\mathrm{sp}(s) + \mathbf{P}_\mathrm{r}(s)$ and*
- *the error measured in the $\mathcal{H}_\infty$ norm satisfies*

$$\|\mathbf{G} - \mathbf{G}_\mathrm{r}\|_{\mathcal{H}_\infty} = \|\mathbf{G}^\mathrm{sp} - \mathbf{G}_\mathrm{r}^\mathrm{sp}\|_{\mathcal{H}_\infty} \leq 2\left(\theta^\mathrm{p}_{q_f+1},\, \dots,\, \theta^\mathrm{p}_{n_f}\right)\,. \tag{5.20}$$

Note that because only zero improper HSVs are truncated in Algorithm 5.1, they do not contribute to the error $\|\mathbf{G} - \mathbf{G}_\mathrm{r}\|_{\mathcal{H}_\infty}$, which leads to the essential equality $\mathbf{P}_\mathrm{r}(s) = \mathbf{P}(s)$. Moreover the realization $[\mathbf{E}_\mathrm{r},\, \mathbf{A}_\mathrm{r},\, \mathbf{B}_\mathrm{r},\, \mathbf{C}_\mathrm{r}]$ is balanced, such that $\mathbf{E}_r$ and $\mathbf{A}_r$ are block-diagonal (see (5.19)). This complies with the idea of splitting the FOM into its slow and fast subsystems.

5.3 Application to the Improper Subsystem of Structured DAEs

In the previous section Lyapunov BT for DAEs has been treated. Although this technique is designed for the reduction of the entire FOM, only the manipulation of the fast subsystem will be used within the scope of this thesis. This way one can exploit the structure of the DAE, to efficiently obtain a minimal realization of the improper subsystem. Furthermore the strictly proper subsystem is adaptively reduced by CUREd SPARK according to Chapter 4, such that BT of the slow subsystem (which requires the solution of large-scale Lyapunov equations) is avoided. This in turn allows the reduction of true large-scale models.

A study of Algorithm 5.1 allows to identify the computation of the Gramians and their Cholesky factors as the bottleneck considering numerical efforts. Even techniques, which directly compute the Cholesky factors of the proper Gramians (instead of the Gramians themselves), are either restricted to small and medium sized problems, or make use of low-rank approximations (such that $\mathbf{\Gamma}^\mathrm{pc} \approx \mathbf{\Omega}^\mathrm{pc}\,\mathbf{\Omega}^{\mathrm{pc}*}$) [9, p. 22f.]. Fortunately in the case of the improper Gramians, an explicit analytic formulation for the Cholesky factors is possible avoiding the computation of the generalized projected discrete-time Lyapunov equations. Using this, the problem of solving large-scale Lyapunov equations is exchanged by solving few sparse linear systems of equations (LSEs), such that the numerical effort is reduced significantly. Furthermore the solution obtained through this method is exact in contrast to using approximate solvers.

Important note: The following results are extracted from [9, p. 22f.] and [39, p. 196f.]. Since a *detailed* proof is missing in the sources, an explicit derivation is presented below. Note that the derivation steps originate from a discussion with the author of [39] during a workshop on MOR held in late June at the Chair of Automatic Control (TUM). Therefore all credits go to Tatjana Stykel from the university of Augsburg.

Key Theorem 5.12 (adapted from [9, p. 22f.] and [39, p. 196f.])**.** *Let* $(\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C},\, \mathbf{x}_0)$ *be an asymptotically stable DAE-system of index* ν *and* $\mathbf{\Pi}_l^\infty$ *and* $\mathbf{\Pi}_r^\infty$ *denote the spectral projectors onto the left and right deflating subspace of* $\lambda\,\mathbf{E} - \mathbf{A}$ *corresponding to the infinite eigenvalues according to Definition 2.8.*

Then the improper controllability and observability Gramians $\mathbf{\Gamma}^{\mathrm{imc}}$ *and* $\mathbf{\Gamma}^{\mathrm{imo}}$ *related to the realization* $[\mathbf{E},\, \mathbf{A},\, \mathbf{B},\, \mathbf{C}]$ *defined in (2.49) and (2.50) satisfy*

$$\mathbf{\Gamma}^{\mathrm{imc}} = \sum_{w=0}^{\nu-1} \left(\mathbf{A}^{-1}\,\mathbf{E}\right)^w \hat{\mathbf{B}}\,\hat{\mathbf{B}}^* \left(\mathbf{E}^*\,\mathbf{A}^{-*}\right)^w \,, \qquad \textit{with} \quad \hat{\mathbf{B}} := \mathbf{\Pi}_r^\infty\,\mathbf{A}^{-1}\,\mathbf{B}\,, \tag{5.21}$$

$$\mathbf{\Gamma}^{\mathrm{imo}} = \sum_{w=0}^{\nu-1} \left(\mathbf{A}^{-*}\,\mathbf{E}^*\right)^w \hat{\mathbf{C}}^*\,\hat{\mathbf{C}} \left(\mathbf{E}\,\mathbf{A}^{-1}\right)^w \,, \qquad \textit{with} \quad \hat{\mathbf{C}} := \mathbf{C}\,\mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty\,, \tag{5.22}$$

while their Cholesky factors $\mathbf{\Omega}^{\mathrm{imc}} \in \mathbb{C}^{n\times\nu\,m}$ *and* $\mathbf{\Omega}^{\mathrm{imo}} \in \mathbb{C}^{n\times\nu\,p}$ *defined in (5.9) read as*

$$\mathbf{\Omega}^{\mathrm{imc}} = \left[\mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty\,\mathbf{B},\, \left(\mathbf{A}^{-1}\,\mathbf{E}\right)\mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty\,\mathbf{B},\, \ldots,\, \left(\mathbf{A}^{-1}\,\mathbf{E}\right)^{\nu-1}\mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty\,\mathbf{B}\right], \tag{5.23}$$

$$\mathbf{\Omega}^{\mathrm{imo}} = \left[\mathbf{A}^{-*}\,\mathbf{\Pi}_r^{\infty *}\,\mathbf{C}^*,\, \left(\mathbf{A}^{-*}\,\mathbf{E}^*\right)\mathbf{A}^{-*}\,\mathbf{\Pi}_r^{\infty *}\,\mathbf{C}^*,\, \ldots,\, \left(\mathbf{A}^{-*}\,\mathbf{E}^*\right)^{\nu-1}\mathbf{A}^{-*}\,\mathbf{\Pi}_r^{\infty *}\,\mathbf{C}^*\right]. \tag{5.24}$$

Proof. At first the regularity of $\mathbf{A}$, thus the existence of $\mathbf{A}^{-1}$, follows from Lemma B.4.

Next consider $\mathbf{\Gamma}^{\mathrm{imc}}$ as the unique solution of the generalized projected discrete-time Lyapunov equation (2.49):

$$\mathbf{A}\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{A}^* - \mathbf{E}\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{E}^* - \mathbf{\Pi}_l^\infty\,\mathbf{B}\,\mathbf{B}^*\,\mathbf{\Pi}_l^{\infty *} = \mathbf{0}\,, \quad \text{with } \mathbf{\Gamma}^{\mathrm{imc}} = \mathbf{\Pi}_r^\infty\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{\Pi}_r^{\infty *}\,. \tag{5.25}$$

Since $\mathbf{A}$ is regular, this is equivalent to

$$\mathbf{\Gamma}^{\mathrm{imc}} - \mathbf{A}^{-1}\,\mathbf{E}\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{E}^*\,\mathbf{A}^{-*} = \mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty\,\mathbf{B}\,\mathbf{B}^*\,\mathbf{\Pi}_l^{\infty *}\,\mathbf{A}^{-*}\,, \tag{5.26}$$

which can be reformulated with the help of Corollary B.5 to

$$\mathbf{\Gamma}^{\mathrm{imc}} - \mathbf{A}^{-1}\,\mathbf{E}\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{E}^*\,\mathbf{A}^{-*} = \underbrace{\mathbf{\Pi}_r^\infty\,\mathbf{A}^{-1}\,\mathbf{B}}_{\hat{\mathbf{B}}}\,\underbrace{\mathbf{B}^*\,\mathbf{A}^{-*}\,\mathbf{\Pi}_r^{\infty *}}_{\hat{\mathbf{B}}^*} = \hat{\mathbf{B}}\,\hat{\mathbf{B}}^*\,. \tag{5.27}$$

Inserting (5.21) into (5.27) leads to

$$\begin{aligned}
&\sum_{w=0}^{\nu-1} \left(\mathbf{A}^{-1}\,\mathbf{E}\right)^w \hat{\mathbf{B}}\,\hat{\mathbf{B}}^* \left(\mathbf{E}^*\,\mathbf{A}^{-*}\right)^w - \sum_{w=1}^{\nu} \left(\mathbf{A}^{-1}\,\mathbf{E}\right)^w \hat{\mathbf{B}}\,\hat{\mathbf{B}}^* \left(\mathbf{E}^*\,\mathbf{A}^{-*}\right)^w = \\
&\qquad = \hat{\mathbf{B}}\,\hat{\mathbf{B}}^* - \left(\mathbf{A}^{-1}\,\mathbf{E}\right)^\nu \mathbf{\Pi}_r^\infty\,\mathbf{A}^{-1}\,\mathbf{B}\,\mathbf{B}^*\,\mathbf{A}^{-*}\,\mathbf{\Pi}_r^{\infty *} \left(\mathbf{E}^*\,\mathbf{A}^{-*}\right)^\nu = \\
&\qquad \overset{(B.16)}{=} \hat{\mathbf{B}}\,\hat{\mathbf{B}}^*
\end{aligned} \tag{5.28}$$

which proves, that (5.21) is indeed a solution of (5.27). Since the condition $\mathbf{\Gamma}^{\mathrm{imc}} = \mathbf{\Pi}_r^\infty\,\mathbf{\Gamma}^{\mathrm{imc}}\,\mathbf{\Pi}_r^{\infty *}$ can be easily verified for (5.21) with (B.15) and $\mathbf{\Pi}_r^\infty\,\hat{\mathbf{B}} = \hat{\mathbf{B}}$, it is shown, that (5.21) is the *unique* solution of (2.49). The proof for (5.22) as the unique solution of (2.50) is obtained in an analogous way.

Finally an explicit computation of the products $\mathbf{\Omega}^{\text{imc}}\,\mathbf{\Omega}^{\text{imc}*}$ and $\mathbf{\Omega}^{\text{imo}}\,\mathbf{\Omega}^{\text{imo}*}$ with (5.23) and (5.24) shows, that $\mathbf{\Omega}^{\text{imc}}$ and $\mathbf{\Omega}^{\text{imo}}$ are the Cholesky factors of $\mathbf{\Gamma}^{\text{imc}}$ and $\mathbf{\Gamma}^{\text{imo}}$, such that $\mathbf{\Gamma}^{\text{imc}} = \mathbf{\Omega}^{\text{imc}}\,\mathbf{\Omega}^{\text{imc}*}$ and $\mathbf{\Gamma}^{\text{imo}} = \mathbf{\Omega}^{\text{imo}}\,\mathbf{\Omega}^{\text{imo}*}$ holds. ■

Note that the special structures of (5.23) and (5.24) allow a description as (projected) block input/output rational Krylov subspaces:

Corollary 5.13. *Let all conditions of Theorem 5.12 hold. Then $\mathbf{\Omega}^{\text{imc}}$ and $\mathbf{\Omega}^{\text{imo}}$ from (5.23) and (5.24) are the primitive bases of the* projected *block input/output rational Krylov subspaces*

$$\begin{aligned} \mathcal{K}^{\infty}_{\text{bi}}\,(\nu) &= \mathcal{K}_{\nu}\left(\mathbf{A}^{-1}\,\mathbf{E},\, \mathbf{A}^{-1}\,\mathbf{\Pi}_l^{\infty}\,\mathbf{B}\right)\ , \\ \mathcal{K}^{\infty}_{\text{bo}}\,(\nu) &= \mathcal{K}_{\nu}\left(\mathbf{A}^{-*}\,\mathbf{E}^{*},\, \mathbf{A}^{-*}\,\mathbf{\Pi}_r^{\infty*}\,\mathbf{C}^{*}\right)\ . \end{aligned} \tag{5.29}$$

This allows an efficient implementation similar to the generation of $\mathbf{V}^{\text{P}}$ in Section 3.2, where instead of $\mathbf{B}$ and $\mathbf{C}$ their projections $\mathbf{\Pi}_l^{\infty}\,\mathbf{B}$ and $\mathbf{C}\,\mathbf{\Pi}_r^{\infty}$ and the special expansion point $s = 0$ are used. Note that in contrast to MOR by rational Krylov subspace methods, where an arbitrary base of $\mathcal{K}_{\text{ti}}$ or $\mathcal{K}_{\text{to}}$ can be used, the factors $\mathbf{\Omega}^{\text{imc}}$ and $\mathbf{\Omega}^{\text{imo}}$ coincide only with the *primitive* base. Therefore an orthogonalization technique as suggested in Section 3.2 may not be applicable in this case.[4]

Apart from that, an additional projection by

$$\mathbf{\Omega}^{\text{imc}} \leftarrow \mathbf{\Pi}_r^{\infty}\,\mathbf{\Omega}^{\text{imc}} \quad \text{and} \quad \mathbf{\Omega}^{\text{imo}} \leftarrow \mathbf{\Pi}_l^{\infty*}\,\mathbf{\Omega}^{\text{imo}} \tag{5.30}$$

does not change the result from an analytic point of view, but may help to avoid numerical issues [39, p. 197].

Finally a simplified version of Algorithm 5.1, which uses the analytic expressions for $\mathbf{\Omega}^{\text{imc}}$ and $\mathbf{\Omega}^{\text{imo}}$ from Theorem 5.12 to process the improper subsystem, is formulated:

Algorithm 5.2 : Computation of a minimal realization of the improper subsystem

```
Input : FOM: [$\mathbf{E}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$], spectral projectors: $\mathbf{\Pi}_l^{\infty}$, $\mathbf{\Pi}_r^{\infty}$ and index $\nu$
Output : minimal realization of improper subsystem: [$\mathbf{E}_{\text{r}}^{\text{im}}$, $\mathbf{A}_{\text{r}}^{\text{im}}$, $\mathbf{B}_{\text{r}}^{\text{im}}$, $\mathbf{C}_{\text{r}}^{\text{im}}$]
compute the Cholesky factors $\mathbf{\Omega}^{\text{imc}}$ and $\mathbf{\Omega}^{\text{imo}}$ according to (5.23) and (5.24)
optional: additional projection of $\mathbf{\Omega}^{\text{imc}}$ and $\mathbf{\Omega}^{\text{imo}}$ according to (5.30)
compute $(\mathbf{Z}_l,\, \mathbf{\Lambda},\, \mathbf{Z}_r) = $ (thin) SVD$\big(\mathbf{\Omega}^{\text{imo}*}\,\mathbf{A}\,\mathbf{\Omega}^{\text{imc}}\big)$        // see Lemma 5.8
set $q_{\infty}$ as the count of non-zero improper HSVs (obtained from $\mathbf{\Lambda}$)
truncate: $\hat{\mathbf{Z}}_l = [\mathbf{z}_{l,1},\, \dots,\, \mathbf{z}_{l,q_{\infty}}]$, $\hat{\mathbf{Z}}_r = [\mathbf{z}_{r,1},\, \dots,\, \mathbf{z}_{r,q_{\infty}}]$, $\hat{\mathbf{\Lambda}} = \text{diag}(\theta_1^{\text{im}},\, \dots,\, \theta_{q_{\infty}}^{\text{im}})$
compute $\mathbf{W}_{\infty} = \mathbf{\Omega}^{\text{imo}}\,\hat{\mathbf{Z}}_l\,(\hat{\mathbf{\Lambda}})^{-1/2}$ and $\mathbf{V}_{\infty} = \mathbf{\Omega}^{\text{imc}}\,\hat{\mathbf{Z}}_r\,(\hat{\mathbf{\Lambda}})^{-1/2}$
project FOM: $\mathbf{E}_{\text{r}}^{\text{im}} = \mathbf{W}_{\infty}^{\text{T}}\,\mathbf{E}\,\mathbf{V}_{\infty}$, $\mathbf{A}_{\text{r}}^{\text{im}} = \mathbf{W}_{\infty}^{\text{T}}\,\mathbf{A}\,\mathbf{V}_{\infty}$, $\mathbf{B}_{\text{r}}^{\text{im}} = \mathbf{W}_{\infty}^{\text{T}}\,\mathbf{B}$ and $\mathbf{C}_{\text{r}}^{\text{im}} = \mathbf{C}\,\mathbf{V}_{\infty}$
```

[4] A definite statement concerning this issue requires further investigations. Note that the order of the rational Krylov subspaces $\mathcal{K}^{\infty}_{\text{bi}}$ and $\mathcal{K}^{\infty}_{\text{bo}}$ is determined by the index of the DAE-system ν. Therefore the column count of $\mathbf{\Omega}^{\text{imc}}$ and $\mathbf{\Omega}^{\text{imo}}$ is rather low in most technical applications (i. e. for index 1 to 3). Thus orthogonalization may not be necessary anyway.

Note that the SVD involved in Algorithm 5.2 is computationally cheap, since the matrix $\mathbf{\Omega}^{\mathrm{imo}*}\,\mathbf{A}\,\mathbf{\Omega}^{\mathrm{imc}} \in \mathbb{C}^{\nu p \times \nu m}$ is of low dimension. Moreover the computation of $\mathbf{A}_{\mathrm{r}}^{\mathrm{im}}$ is unnecessary, since

$$\begin{aligned}\mathbf{A}_{\mathrm{r}}^{\mathrm{im}} &= \mathbf{W}_{\infty}^{*}\,\mathbf{A}\,\mathbf{V}_{\infty} = (\hat{\mathbf{\Lambda}})^{-1/2}\,\hat{\mathbf{Z}}_l^{*}\,\underbrace{\mathbf{\Omega}^{\mathrm{imo}*}\,\mathbf{A}\,\mathbf{\Omega}^{\mathrm{imc}}}_{=\hat{\mathbf{Z}}_l\,\hat{\mathbf{\Lambda}}\,\hat{\mathbf{Z}}_r^{*}\ \text{(SVD)}}\,\hat{\mathbf{Z}}_r\,(\hat{\mathbf{\Lambda}})^{-1/2}\\ &= (\hat{\mathbf{\Lambda}})^{-1/2}\,\hat{\mathbf{Z}}_l^{*}\,\hat{\mathbf{Z}}_l\,\hat{\mathbf{\Lambda}}\,\hat{\mathbf{Z}}_r^{*}\,\hat{\mathbf{Z}}_r\,(\hat{\mathbf{\Lambda}})^{-1/2}\\ &= \ldots \text{unitarity of } \hat{\mathbf{Z}}_l \text{ and } \hat{\mathbf{Z}}_r \ldots = (\hat{\mathbf{\Lambda}})^{-1/2}\,\hat{\mathbf{\Lambda}}\,(\hat{\mathbf{\Lambda}})^{-1/2} = \mathbf{I}_{q_\infty}\end{aligned} \tag{5.31}$$

holds. Herein $\mathbf{Z}_l\,\mathbf{\Lambda}\,\mathbf{Z}_r^{*} = \hat{\mathbf{Z}}_l\,\hat{\mathbf{\Lambda}}\,\hat{\mathbf{Z}}_r^{*}$ has been used.

Similar to Theorem 5.11, the preconditions and properties of Algorithm 5.2 are summarized in Corollary 5.14:

Corollary 5.14 (derived from Theorem 5.11)**.** *Let*

- *the FOM be described by the asymptotically stable DAE-system* $(\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C},\,\mathbf{x}_0)$ *of index* ν *with transfer function* $\mathbf{G}(s)$*,*
- $\mathbf{\Pi}_l^{\infty}$ *and* $\mathbf{\Pi}_r^{\infty}$ *denote spectral projectors according to Definition 2.8 which are known in advance and*
- $\mathbf{G}(s)$ *be composed of* $\mathbf{G}^{\mathrm{sp}}(s)$ *and* $\mathbf{P}(s)$ *corresponding to the strictly proper and improper subsystem respectively, i. e.* $\mathbf{G}(s) = \mathbf{G}^{\mathrm{sp}}(s) + \mathbf{P}(s)$*.*

Then $[\mathbf{E}_{\mathrm{r}}^{\mathrm{im}},\,\mathbf{A}_{\mathrm{r}}^{\mathrm{im}},\,\mathbf{B}_{\mathrm{r}}^{\mathrm{im}},\,\mathbf{C}_{\mathrm{r}}^{\mathrm{im}}]$ *obtained by Algorithm 5.2 is a minimal realization of* $\mathbf{P}(s)$*.*

Remark 5.15*.* As shown in Theorem 5.12, the analytic expressions for the Cholesky factors involve the DAE-index ν. Note that although this information may be available in most applications (since structured problems are considered), the index of the FOM has not to be known in advance. This is because of Corollary B.5 which states

$$\left(\mathbf{A}^{-1}\,\mathbf{E}\right)^{\nu}\,\mathbf{\Pi}_r^{\infty} = \mathbf{0} \quad \text{and} \quad \left(\mathbf{A}^{-*}\,\mathbf{E}^{*}\right)^{\nu}\,\mathbf{\Pi}_l^{\infty*} = \mathbf{0} \tag{5.32}$$

or equivalently

$$\left(\mathbf{A}^{-1}\,\mathbf{E}\right)^{\nu}\,\underbrace{\mathbf{A}^{-1}\mathbf{\Pi}_l^{\infty}}_{\mathbf{\Pi}_r^{\infty}\,\mathbf{A}^{-1}}\,\mathbf{B} = \mathbf{0} \quad \text{and} \quad \left(\mathbf{A}^{-*}\,\mathbf{E}^{*}\right)^{\nu}\,\underbrace{\mathbf{A}^{-*}\,\mathbf{\Pi}_r^{\infty*}}_{\mathbf{\Pi}_l^{\infty*}\,\mathbf{A}^{-*}}\,\mathbf{C}^{*} = \mathbf{0}\,. \tag{5.33}$$

Thus the blocks of $\mathbf{\Omega}^{\mathrm{imc}}$ and $\mathbf{\Omega}^{\mathrm{imo}}$ can be computed recursively according to (5.23) and (5.24), until a zero block occurs which finishes the recursion. A possible implementation is given in [39, p. 197] (therein denoted as “The generalized Smith method for the projected GDALE”). Be aware that this strategy does not allow to determine the index numerically, since the blocks of $\mathbf{\Omega}^{\mathrm{imc}}$ and $\mathbf{\Omega}^{\mathrm{imo}}$ may become zero independently of the index (imagine the case $\mathbf{\Pi}_l^{\infty}\,\mathbf{B} = \mathbf{0}$ or $\mathbf{C}\,\mathbf{\Pi}_r^{\infty} = \mathbf{0}$).

Chapter 6

Summary of the Main Results

In the previous chapters an adaptive scheme for the reduction of structured, improper DAE-systems has been presented. Since the derivation is rather lengthy, one may lose track of the core statements. To avoid this, a compact summary of the main results of this thesis is given in the following.

First an overview of the investigated algorithms is given in Section 6.1. Furthermore the overall procedure connecting the results from Chapter 4 and Chapter 5 is formulated as pseudo-code. The main focus is set on the presentation of the requirements regarding the FOM and the properties of the resulting ROM. Thus this section is suitable as a quick reference (e. g. during implementation).

Finally Section 6.2 contains a recapitulation of the $\mathcal{H}_2$ inner-product of two strictly proper DAEs (or more precisely: of their transfer functions) as derived in Section 4.1. Although this result has been used to adapt the PORK algorithm to the DAE-case and thus is part of the derivation, it is treated separately in the following. This is because it describes a very general relationship which can be used for independent investigations which may be not related to MOR at all.

6.1 Adaptive $\mathcal{H}_2$ Pseudo-Optimal Reduction of Improper DAEs

This section contains an overview of the presented algorithms of Chapter 4 and Chapter 5. As tangential-input rational Krylov subspaces have been in the focus during derivation, only the corresponding "**V**-based" versions are considered in the following, while the "**W**-based" algorithms can be obtained using the duality principle in linear systems. Certainly analogous requirements and properties hold in the dual case.

First of all it has been shown in Section 4.2 that the $\mathcal{H}_2$ pseudo-optimal rational Krylov (PORK) algorithm as formulated in [42, p. 91] is directly (i. e. without any modifications) applicable to the case of asymptotically stable and *strictly proper* DAEs. Note that the restriction to strictly proper systems is essential, since the $\mathcal{H}_2$ inner-product and thus the concept of $\mathcal{H}_2$ pseudo-optimality is not defined for proper or improper transfer functions. As in the ODE-case additional assumptions (e. g. the existence of a realization with real-valued system matrices) have to be made, which are easily satisfied in most technical applications.

Beside the rather specific PORK algorithm, the general conditions for $\mathcal{H}_2$ pseudo-optimality given in [42, p. 87] have been investigated in the context of DAEs during the work on this thesis. Almost all conditions can be transferred to the DAE-case, the sole exception has been discussed at the end of Section 4.2.

Furthermore the stability-preserving, adaptive rational Krylov (SPARK) algorithm and the cumulative reduction (CURE) framework as presented in [30, p. 75ff.] and [42, p. 49ff.] have been verified for the DAE-case in Section 4.3. Since these techniques are related to the PORK algorithm (at least within this work), they inherit its requirements. It is worth noting, that neither SPARK nor CURE introduce additional *DAE-related* requirements, thus they can be applied “out of the box”. Like in the ODE-case SPARK (currently) only works with SISO-systems, while CURE requires the matrix $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ (and further $[\mathbf{E}\,\hat{\mathbf{V}},\, \mathbf{B}_\perp]$ in each iteration) to be of full rank.

Finally a minimal realization of the improper subsystem has been derived in Chapter 5. This is done by BT as summarized in [9] whose main requirement is asymptotic stability of the FOM. As this property is necessary for the PORK algorithm as well, no additional conditions are introduced.

The general framework for the reduction of improper DAEs has been presented in Section 3.5 and is based on the partitioning of the FOM into a strictly proper (slow) and improper (fast) subsystem using the spectral projectors. The strictly proper subsystem is reduced with CUREd SPARK which guarantees stability as well as $\mathcal{H}_2$ pseudo-optimality and allows to specify the order of the ROM (or more precisely: of its strictly proper contribution). After that the polynomial part of the original transfer function is incorporated by the minimal realization derived in Chapter 5. The overall procedure is described in Algorithm 6.1:

Algorithm 6.1 : Adaptive MOR of structured, improper SISO-DAEs

```
Input : FOM: $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}]$, spectral projectors $\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_r^f$ and index $\nu$
Output : ROM: $[\mathbf{E}_\mathrm{r}, \mathbf{A}_\mathrm{r}, \mathbf{B}_\mathrm{r}, \mathbf{C}_\mathrm{r}]$

// separation of the strictly proper and improper subsystem
$\mathbf{C}^\mathrm{sp} = \mathbf{C}\,\mathbf{\Pi}_r^f$        // strictly proper subsystem: $[\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}^\mathrm{sp}]$
$\mathbf{B}^\mathrm{im} = \mathbf{\Pi}_l^\infty\,\mathbf{B} = \mathbf{B} - \mathbf{\Pi}_l^f\,\mathbf{B}$
$\mathbf{C}^\mathrm{im} = \mathbf{C}\,\mathbf{\Pi}_r^\infty = \mathbf{C} - \mathbf{C}^\mathrm{sp}$        // improper subsystem: $[\mathbf{E}, \mathbf{A}, \mathbf{B}^\mathrm{im}, \mathbf{C}^\mathrm{im}]$

// reduction of the strictly proper subsystem with CUREd SPARK
$(\mathbf{E}_\mathrm{r}^\mathrm{sp}, \mathbf{A}_\mathrm{r}^\mathrm{sp}, \mathbf{B}_\mathrm{r}^\mathrm{sp}, \mathbf{C}_\mathrm{r}^\mathrm{sp}) = \mathrm{CUREd\ SPARK}(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}^\mathrm{sp})$        // see Algorithm 4.4

// compute minimal realization of the improper subsystem with BT for DAEs
$(\mathbf{E}_\mathrm{r}^\mathrm{im}, \mathbf{A}_\mathrm{r}^\mathrm{im}, \mathbf{B}_\mathrm{r}^\mathrm{im}, \mathbf{C}_\mathrm{r}^\mathrm{im}) = \mathrm{DAE\text{-}BT}(\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{B}^\mathrm{im}, \mathbf{C}^\mathrm{im}, \nu)$[1]        // see Algorithm 5.2

// (re)connection of the subsystems to the final ROM
$\mathbf{E}_\mathrm{r} = \begin{bmatrix} \mathbf{E}_\mathrm{r}^\mathrm{sp} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_\mathrm{r}^\mathrm{im} \end{bmatrix}$, $\mathbf{A}_\mathrm{r} = \begin{bmatrix} \mathbf{A}_\mathrm{r}^\mathrm{sp} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_\mathrm{r}^\mathrm{im} \end{bmatrix}$, $\mathbf{B}_\mathrm{r} = \begin{bmatrix} \mathbf{B}_\mathrm{r}^\mathrm{sp} \\ \mathbf{B}_\mathrm{r}^\mathrm{im} \end{bmatrix}$ and $\mathbf{C}_\mathrm{r} = \begin{bmatrix} \mathbf{C}_\mathrm{r}^\mathrm{sp} & \mathbf{C}_\mathrm{r}^\mathrm{im} \end{bmatrix}$
```

[1] Instead of the spectral projectors $\mathbf{\Pi}_l^\infty$ and $\mathbf{\Pi}_r^\infty$ (which are dense, large-scale matrices), the projected input and output matrices $\mathbf{B}^\mathrm{im}$ and $\mathbf{C}^\mathrm{im}$ are passed to Algorithm 5.2. This is an implementation detail to reduce computation time and memory consumption. Note that the computation of $\mathbf{\Omega}^\mathrm{imc}$ and $\mathbf{\Omega}^\mathrm{imo}$ in Algorithm 5.2 does not require the explicit knowledge of the spectral projectors, instead the products $\mathbf{\Pi}_l^\infty\,\mathbf{B} = \mathbf{B}^\mathrm{im}$ and $\mathbf{C}\,\mathbf{\Pi}_r^\infty = \mathbf{C}^\mathrm{im}$ are sufficient.

As stated in Section 5.3, the index ν of the DAE does not have to be known in advance. Furthermore all theoretical results apply to general (linear) DAEs of arbitrary structure, index and properness. However, analytic expressions of the spectral projectors ($\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_r^f$ or $\mathbf{\Pi}_l^\infty$, $\mathbf{\Pi}_r^\infty$) are necessary in order to allow an efficient implementation.

Finally Figure 6.1 gives a complete view on all requirements and properties of the presented algorithms.

		Requirements regarding the FOM							*Properties of the ROM*			
Algorithm		$\lambda\mathbf{E}-\mathbf{A}$ regular	consistent initial value	asymptotic stability	strictly properness	allows realization with real-valued system matrices	SISO	$[\mathbf{E}\,\mathbf{V},\mathbf{B}]$ has full column rank	asymptotic stability	tangential interpol.	$\mathcal{H}_2$ pseudo-optimality	local $\mathcal{H}_2$ optimality
4.1	PORK	●	●	●	●	●			●	●	●	
4.3	SPARK	●	●	●	●	●	●		●	●	●	●
4.4	CUREd SPARK	●	●	●	●	●	●	●	●	○	●	
5.2	DAE-BT of $\mathbf{P}(s)$	●	●	●					●			
6.1	overall procedure	●	●	●		●	●	●	●	○	●	

Figure 6.1: Requirements of the presented algorithms and properties of the resulting ROM. The symbol ● indicates, that either a requirement has to be met or a property of the ROM is guaranteed.

Note the symbol ○ in Figure 6.1 which indicates, that tangential interpolation is achieved "halfway" during the CURE-framework. In particular the expansion points chosen in each iteration are preserved, while the tangential directions are transformed during the assembly to the overall ROM [42, p. 60f.]. As the combination with SPARK is restricted to SISO-systems, the issue of tangential directions is irrelevant. Thus a big advantage of CUREd SPARK is, that new expansion points are added during the CURE-iteration without destroying previously incorporated interpolation data [42, p. 61].

Furthermore note that the overall ROM $\mathbf{G}_\mathrm{r}(s) = \mathbf{G}_\mathrm{r}^\mathrm{sp}(s) + \mathbf{P}_\mathrm{r}(s)$ as well as its strictly proper part $\mathbf{G}_\mathrm{r}^\mathrm{sp}(s)$ are $\mathcal{H}_2$ pseudo-optimal (with respect to the subspace of transfer functions $\mathcal{G}$ defined by the set of expansion points). This is because $\mathbf{G}_\mathrm{r}^\mathrm{sp}(s)$ is obtained through applying CUREd SPARK onto $\mathbf{G}^\mathrm{sp}$ which leads to

$$\mathbf{G}_\mathrm{r}^\mathrm{sp}(s) = \arg \min_{\hat{\mathbf{G}}^\mathrm{sp} \in \mathcal{G}} \|\mathbf{G}^\mathrm{sp} - \hat{\mathbf{G}}^\mathrm{sp}\|_{\mathcal{H}_2} \,. \tag{6.1}$$

Furthermore Algorithm 5.2 guarantees $\mathbf{P}_\mathrm{r}(s) = \mathbf{P}(s)$, such that

$$\|\mathbf{G} - \hat{\mathbf{G}}\|_{\mathcal{H}_2} = \|\mathbf{G}^\mathrm{sp} + \mathbf{P} - \hat{\mathbf{G}}^\mathrm{sp} - \mathbf{P}\|_{\mathcal{H}_2} = \|\mathbf{G}^\mathrm{sp} - \hat{\mathbf{G}}^\mathrm{sp}\|_{\mathcal{H}_2} \tag{6.2}$$

with $\hat{\mathbf{G}}(s) := \hat{\mathbf{G}}^\mathrm{sp}(s) + \mathbf{P}(s)$ and $\hat{\mathbf{G}}^\mathrm{sp} \in \mathcal{G}$ holds. This finally proves

$$\arg \min_{\hat{\mathbf{G}}} \|\mathbf{G} - \hat{\mathbf{G}}\|_{\mathcal{H}_2} = \arg \min_{\hat{\mathbf{G}}^\mathrm{sp} \in \mathcal{G}} \|\mathbf{G}^\mathrm{sp} - \hat{\mathbf{G}}^\mathrm{sp}\|_{\mathcal{H}_2} + \mathbf{P}(s) = \mathbf{G}_\mathrm{r}^\mathrm{sp}(s) + \mathbf{P}(s) = \mathbf{G}_\mathrm{r}(s) \,. \tag{6.3}$$

6.2 $\mathcal{H}_2$ Inner-Product of Strictly Proper DAEs

In Section 4.1 the $\mathcal{H}_2$ inner-product of the transfer functions of two DAE-systems has been derived. Although the inner-product of $\mathcal{H}_2$ functions is a basic result of functional analysis, the formulation in the DAE-context via (projected) generalized Sylvester equation as in Theorem 4.9 seems to be new.

It is important to note that the presented result is restricted to a special type of DAEs. In particular both (LTI-) DAE-systems have to

- be described by a regular matrix pencil $\lambda\mathbf{E} - \mathbf{A}$ and a consistent initial value,
- be asymptotically stable and strictly proper,
- be of the same dimension, i. e. share the same count of inputs and outputs ($m = m_{\mathrm{H}}$ and $p = p_{\mathrm{H}}$, but not necessarily $m = p$ or $m_{\mathrm{H}} = p_{\mathrm{H}}$), and
- allow realizations with real-valued system matrices.

The first three requirements are necessary to formulate a $\mathcal{H}_2$ inner-product at all, while the last one takes care of commutativity (which is exploited during the derivation of Theorem 4.9).

Since Theorem 4.9 describes the most general relation, one can derive several special cases depending on the type of the involved systems (see Table 6.1).

Table 6.1: Cases of Theorem 4.9 depending on the type of the considered systems. Note that the term “equal” is related to the corresponding *realizations* of the systems.

Case	Meaning/Usage
two (different) DAEs	most general case
two equal DAEs	$\mathbf{X}$ and $\mathbf{Y}$ coincide with the *proper* controllability and observability Gramians
one ODE, one DAE	used in Section 4.2 to prove $\mathcal{H}_2$ pseudo-optimality
two (different) ODEs	corresponds to the result in [42, p. 65]
two equal ODEs	$\mathbf{X}$ and $\mathbf{Y}$ coincide with the “usual” controllability and observability Gramians

Although the case of one ODE and one DAE has been in the focus of this thesis, the general result involving two (different) DAEs may be useful in a different context.

Chapter 7

Numerical Examples

While the previous chapters presented a detailed mathematical discussion on adaptive MOR of structured DAE-systems, the actual application of the proposed method (i. e. Algorithm 6.1) is examined by means of several numerical examples in the following.

For this purpose the sparse state-space and model order reduction (sssMOR) toolbox ([10], version 1.05 - May 9 2016) developed at the Chair of Automatic Control (TUM) is used. The toolbox runs in MATLAB (The MathWorks, Inc., `www.mathworks.com`) and extends its functionality by common and state of the art MOR-algorithms for small-, medium- and large-scale systems. All numerical results presented in this thesis (including plots contained in the previous chapters) have been generated with MATLAB R2016a 64bit on Ubuntu 16.04 LTS 64bit. The used hardware involves an AMD Phenom™ II X4 940 CPU together with 8 GB DDR2 system memory. The machine precision is limited to $\varepsilon = 2.22 \cdot 10^{-16}$ (double precision).

Since PORK, SPARK as well as CUREd SPARK are provided by the sssMOR toolbox, only Algorithm 5.2 (for processing the improper subsystem with Lyapunov BT) and Algorithm 6.1 (overall procedure) have been implemented. Furthermore a check of the column rank of $[\mathbf{E}\,\mathbf{V}, \mathbf{B}]$ has been added to the CUREd SPARK algorithm (according to Algorithm 4.4). Note that within all performed benchmarks, this condition has been fulfilled at all times. Thus the considerations concerning $[\mathbf{E}\,\mathbf{V}, \mathbf{B}]$ in Appendix C seem to be mainly of theoretical importance.

In the following sections several numerical examples demonstrating the reduction of *structured* DAEs are presented. Since the focus is set on adaptive MOR with CUREd SPARK, only SISO-systems are considered. While Sections 7.1 to 7.3 treat physically based systems selected from the survey given in [9], an entirely “artificial” system demonstrating the validity even for high-index DAEs is discussed in Section 7.4. Table 7.1 gives an overview of the considered benchmark systems and their characteristics.

Table 7.1: Properties of the investigated benchmark systems.

Structure	Section	Index ν	Dimension n	$\mathcal{O}(\mathbf{P}(s))$	Properness
semi-explicit	7.1	1	13250	0	proper
Stokes-like	7.2	2	19039	–	strictly proper
mechanical	7.3	3	2001	1	improper
artificial	7.4	10	5000	5	improper

7.1 Semi-Explicit Index 1 System

First of all a semi-explicit index 1 DAE-system is considered. This type of systems typically arises in computational fluid dynamics and power systems modeling and is structured as follows [9, p. 28]:

$$\begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} \mathbf{u}(t) \,, \qquad \mathbf{y}(t) = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} \,. \tag{7.1}$$

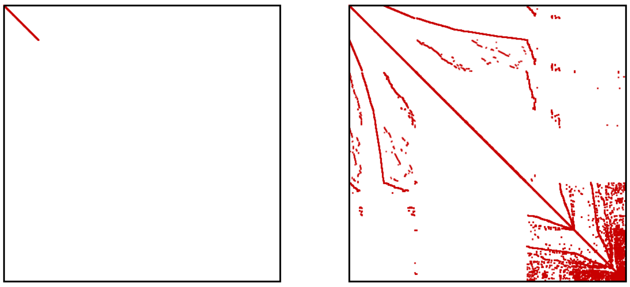
Herein the matrices $\mathbf{E}_{11}$ and $\mathbf{A}_{22} - \mathbf{A}_{21}\,\mathbf{E}_{11}^{-1}\,\mathbf{E}_{12}$ are assumed to be regular, such that $\nu = 1$ holds and analytic expressions for the spectral projectors can be formulated (see [9, p. 28]).

In the following the particular model "BIPS/97" (MIMO46) created by the Brazilian Electrical Energy Research Center (CEPEL) which is available online from the MOR Wiki [33] is investigated. The model describes a power system and is intended for small-signal studies (especially stability analysis and controller design) [33]. It connects $n = 13250$ state variables, where 1664 belong to the so called *dynamical subsystem* (i. e. $n_{\text{dyn}} = \dim(\mathbf{E}_{11}) = 1664$). In contrast to most other benchmark models from [33], this system is improper (or more precisely: proper since $\mathbf{P}(s) = \mathbf{P} = \text{const.}$).

As the original system available from [33] is of MIMO-type ($m = 46$, $p = 46$) only the channel $G_{42,42}(s)$, i. e. $u_{42} \rightarrow y_{42}$ is considered. Furthermore several row- and column-swapping transformations have been applied in order to obtain a structure as in (7.1). Note that the selected model represents a simplification of the general case of semi-explicit index 1 DAE-systems, since $\mathbf{E}_{11} = \mathbf{I}_{n_{\text{dyn}}}$ and $\mathbf{E}_{12} = \mathbf{0}$ holds. This allows to simplify the expressions for the spectral projectors from [9, p. 28] to

$$\mathbf{\Pi}_l^f = \begin{bmatrix} \mathbf{I}_{n_{\text{dyn}}} & -\mathbf{A}_{12}\,\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} , \qquad \mathbf{\Pi}_r^f = \begin{bmatrix} \mathbf{I}_{n_{\text{dyn}}} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\,\mathbf{A}_{21} & \mathbf{0} \end{bmatrix} . \tag{7.2}$$

Although the dimension of the FOM is quite large, the matrices $\mathbf{E}$ and $\mathbf{A}$ are sparse (see Figure 7.1) such that the memory limits of the available hardware are respected.



(a) Sparsity pattern of $\mathbf{E}$

(b) Sparsity pattern of $\mathbf{A}$

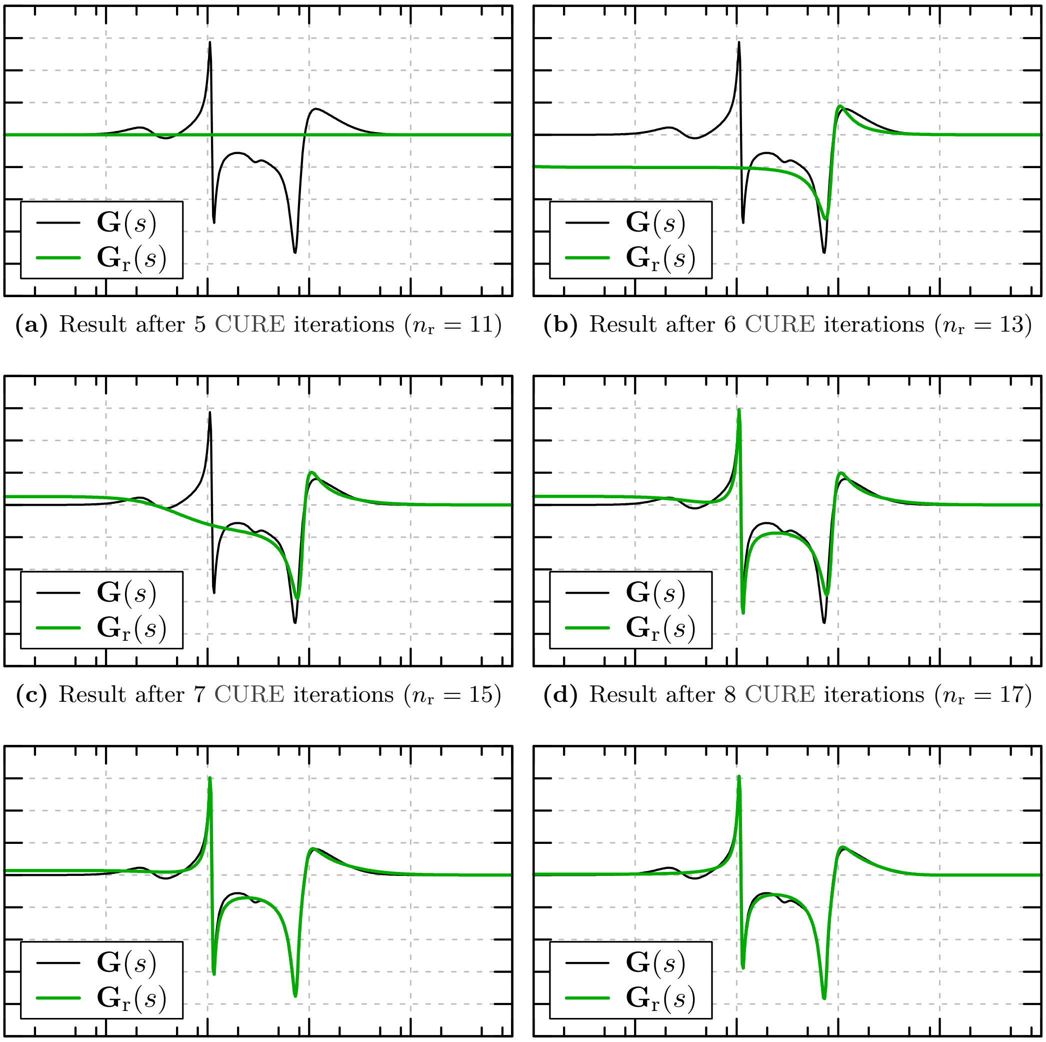
Figure 7.1: Sparsity pattern of the matrices $\mathbf{E}$ and $\mathbf{A}$ of the BIPS/97 (MIMO46) power-system model after reordering[1]. Non-zero entries are indicated by red points.

[1] It is worth noting that modeling techniques like modified nodal analysis often do not provide the FOM in the form of (7.1) directly. Instead row- and column-swapping transformations have to be applied.

In Figure 7.2 several ROMs of different size obtained through the application of Algorithm 6.1 onto the BIPS/97 (MIMO46) benchmark model are presented. Because in each CURE iteration the count of reduced state variables grows by 2 (see Section 4.3), the final dimension of the ROM is given by

$$n_{\mathrm{r}} = \dim(\mathbf{A}_{\mathrm{r}}) = 2 \cdot (\text{count of CURE iterations}) + \dim(\mathbf{A}_{\mathrm{r}}^{\mathrm{im}}) \;, \tag{7.3}$$

where $\dim(\mathbf{A}_{\mathrm{r}}^{\mathrm{im}}) = 1$ (proper, index 1, SISO-DAE). Since each frequency response plot corresponds to a different count of CURE iterations, Figure 7.2 demonstrates the cumulative reduction scheme. Note that the first *visible* approximation of the dynamical subsystem appears after 6 CURE iterations (see Figure 7.2b). This is due to the primitive initialization of the SPARK algorithm with $a = b = 10^{-4}$. Finding appropriate initial values for a and b is a field of research on its own and is not treated within this work.



(a) Result after 5 CURE iterations ($n_{\mathrm{r}} = 11$) **(b)** Result after 6 CURE iterations ($n_{\mathrm{r}} = 13$)

(c) Result after 7 CURE iterations ($n_{\mathrm{r}} = 15$) **(d)** Result after 8 CURE iterations ($n_{\mathrm{r}} = 17$)

(e) Result after 9 CURE iterations ($n_{\mathrm{r}} = 19$) **(f)** Result after 10 CURE iterations ($n_{\mathrm{r}} = 21$)

Figure 7.2: Stepwise reduction of the BIPS/97 (MIMO46) benchmark model with CUREd SPARK (strictly proper subsystem) and Lyapunov BT (improper subsystem). In (a) to (f) the frequency responses (magnitude over frequency) for different target dimensions n_{r} are depicted.

Keep in mind that the processing of the improper subsystem is decoupled from the actual reduction procedure, thus $\mathbf{P}(s)$ (visible as constant feedthrough in Figure 7.2) is matched perfectly at any time. The frequency responses of the resulting error systems $\mathbf{G}_\mathrm{e}(s) = \mathbf{G}(s) - \mathbf{G}_\mathrm{r}(s)$ is depicted in Figure 7.3 which shows the decrease of the error within each iteration.

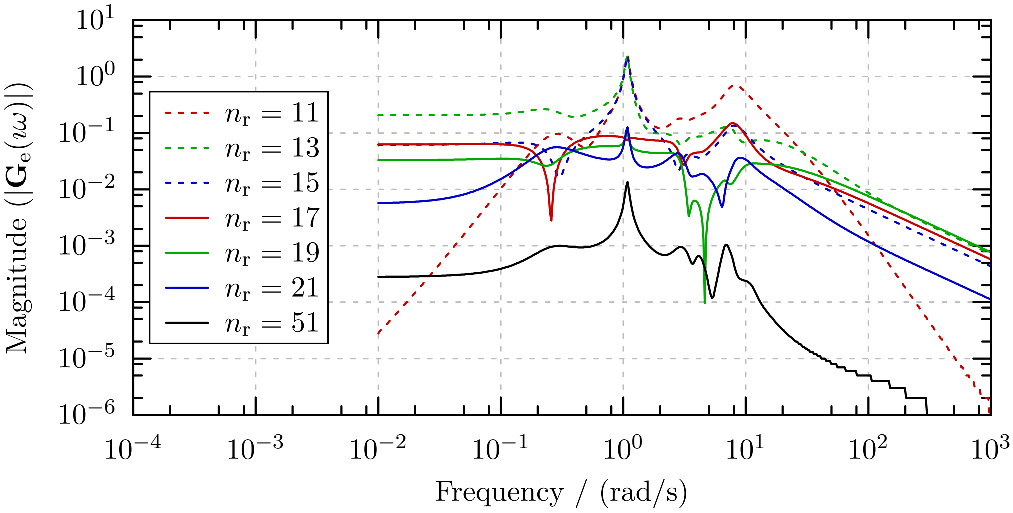


Figure 7.3: Frequency response of the error systems $\mathbf{G}_\mathrm{e}(s) = \mathbf{G}(s) - \mathbf{G}_\mathrm{r}(s)$ corresponding to the results shown in Figures 7.2 and 7.4. With each CURE iteration the (overall) error measured in the $\mathcal{H}_2$ norm decreases (see Theorem 4.54)[2].

Figure 7.4 finally shows the frequency response of the ROM after 25 CURE iterations. Within this example the FOM consists mainly of algebraic equations ($\frac{n-n_\mathrm{dyn}}{n} \approx 87\%$). Because those are simplified to a single constraint in the ROM, an approximation with $n_\mathrm{r} = 51$ as shown in Figure 7.4 appears to be sufficient.

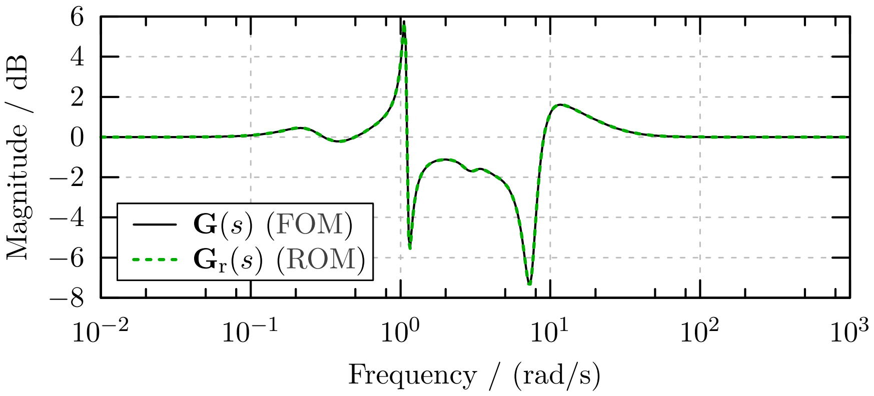


Figure 7.4: Approximation of the FOM of dimension $n = 13250$ with a ROM of dimension $n_\mathrm{r} = 51$ (in 25 CURE iterations). Due to Algorithm 5.2 the 11586 algebraic equations of the FOM are simplified to a single constraint in the ROM.

Note that the same class of DAEs (and even the same benchmark model) has been analyzed in [11]. A comparison of the MOR-techniques shows that there are strong

[2] Note that due to hardware limits, the actual $\mathcal{H}_2$ norm of the error system could not be evaluated.

relations: while the results in [11] are obtained by exploiting the special structure of (7.1) (with $\mathbf{E}_{12} = \mathbf{0}$), the investigations of this thesis are based on a general formulation using the spectral projectors. As a direct comparison of the partitioning by $\mathbf{B}^{\text{sp}} = \mathbf{\Pi}_l^f\,\mathbf{B}$ and $\mathbf{C}^{\text{sp}} = \mathbf{C}\,\mathbf{\Pi}_r^f$ with [11, proposition 2] shows, both methods are equivalent. Anyway the formulation by spectral projectors is not restricted to structures as in (7.1), such that the presented framework represents a more general approach.

Furthermore note that the MOR-technique proposed in [11] is based on the (analytic) identification of the improper part $\mathbf{P}(s)$ (therein called *implicit* feed-through D_{imp}). Since the ROM in [11] incorporates the improper subsystem by appending D_{imp} to the "original" feedthrough D, a slightly smaller dimension of the ROM (in fact by 1) is obtained in comparison to a concatenation of the subsystems according (3.59). Note that since $\nu = 1$ and thus $\mathbf{N}^\nu = \mathbf{N} = \mathbf{0} \Rightarrow \mathbf{E}_{\text{r}}^{\text{im}} = \mathbf{0}$ holds, the same result can be achieved with the strategy presented in Section 3.5. This is done by removing the (reduced) improper subsystem from (3.59). Instead the "implicit" feedthrough $\mathbf{D}_{\text{r}} = \mathbf{C}_{\text{r}}^{\text{im}}\,(\mathbf{A}_{\text{r}}^{\text{im}})^{-1}\mathbf{B}_{\text{r}}^{\text{im}}$ has to be added to the reduced transfer function $\mathbf{G}_{\text{r}}(s)$.

7.2 Stokes-Like Index 2 System

Within this section the reduction of a Stokes-like DAE-system of index 2 is analyzed. Such systems arise in computational fluid dynamics where the flow of an incompressible fluid is modeled by the Navier-*Stokes* equation [9, p. 32]. Linearization and discretization in space by the finite element method leads to a system of the structure [9, p. 32]:

$$\begin{aligned}
\begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} \mathbf{u}(t)\,, \\
\mathbf{y}(t) &= \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix}\,.
\end{aligned} \tag{7.4}$$

Herein $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ typically denote velocity and pressure vectors respectively. If the matrices $\mathbf{E}_{11}$ and $\mathbf{A}_{21}\,\mathbf{E}_{11}^{-1}\,\mathbf{A}_{12}$ are both nonsingular, then the DAE is of index 2 and again analytic expressions for the spectral projectors can be found (see [9, p. 32]).

The actual model used for reduction is generated through a MATLAB script created by Michael Schmidt (at the Technische Universität Berlin in May 2007) and modified by Tatjana Stykel (in November 2007). This script produces a semidiscretized (2D-) Stokes-like system according to [35, p. 34ff.], which has the structure of (7.4) with additionally $\mathbf{E}_{11} = \mathbf{I}_{n_{\text{dyn}}}$. Thus the spectral projectors given in [9, p. 32] simplify to

$$\mathbf{\Pi}_l^f = \begin{bmatrix} \mathbf{K} & -\mathbf{K}\,\mathbf{A}_{11}\,\mathbf{A}_{12}\,(\mathbf{A}_{21}\,\mathbf{A}_{12})^{-1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\,, \quad \mathbf{\Pi}_r^f = \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ -\,(\mathbf{A}_{21}\,\mathbf{A}_{12})^{-1}\,\mathbf{A}_{21}\,\mathbf{A}_{11}\,\mathbf{K} & \mathbf{0} \end{bmatrix}\,, \tag{7.5}$$

with $\mathbf{K} := \mathbf{I}_{n_{\text{dyn}}} - \mathbf{A}_{12}\,(\mathbf{A}_{21}\,\mathbf{A}_{12})^{-1}\,\mathbf{A}_{21}$. Note that the MATLAB script provides explicit (dense) matrices for $\mathbf{\Pi}_l^f$ and $\mathbf{\Pi}_r^f$ which are not used in the following. Instead sparse matrix-vector operations using (7.5) are performed, which significantly reduces memory consumption.

Choosing $m = p = 1$ (SISO) and a 80×80 square grid for spacial discretization leads to a FOM of dimension $n = 19039$ and *dynamical order* $n_{\text{dyn}} = \dim(\mathbf{E}_{11}) = 12640$. The structure of the resulting system matrices $\mathbf{E}$ and $\mathbf{A}$ is depicted in Figure 7.5.

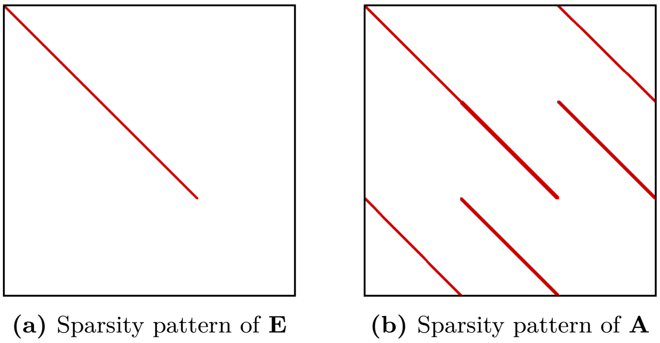


Figure 7.5: Sparsity pattern of the matrices **E** and **A** of the semidiscretized 2D Stokes-like system according to [35, p. 34ff.].

Figure 7.6 shows the reduction result after applying Algorithm 6.1 with one CURE iteration (corresponds to "pure" PORK + SPARK). As the FOM is strictly proper, there is no reduced improper subsystem, i. e. the resulting ROM is of ODE-type. Since $\mathbf{P}(s) = \mathbf{P}_{\mathrm{r}}(s) = 0$ holds, the transfer function may not be matched exactly at $s \to \infty$.

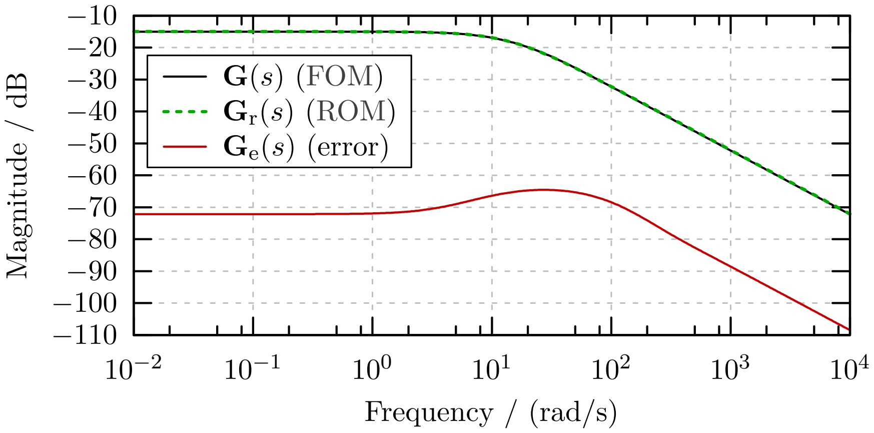


Figure 7.6: Approximation of the FOM $\mathbf{G}(s)$ (Stokes-like index 2 DAE, $n = 19039$) with a ROM $\mathbf{G}_{\mathrm{r}}(s)$ of order $n_{\mathrm{r}} = 2$. Since the FOM involves "simple" dynamics, a low-dimensional ROM is sufficient for small error ($\mathbf{G}_{\mathrm{e}}(s) = \mathbf{G}(s) - \mathbf{G}_{\mathrm{r}}(s)$).

Note that within this example the reduction with Algorithm 6.1 is mathematically equivalent to "usual" ODE-CUREd SPARK as presented in [42] and [30]. This is because strictly properness of the FOM is sufficient for the application of CUREd SPARK (see Figure 6.1). Anyway the projection of **C** with $\mathbf{\Pi}_r^f$ in Algorithm 6.1 might be beneficial from a numerical point of view.

7.3 Mechanical Index 3 System

As third physically based example, a mechanical (multibody) system of index 3 is considered. In particular a constrained damped mass-spring system similar to [28, p. 106] is analyzed (see Figure 7.7). In contrast to [28] the holonomic constraint connecting the first and last mass is removed. Moreover the input $\mathbf{u}(t)$ now directly controls the position of the first mass (instead of applying a force to it), which represents an additional constraint. Note that this modification does not affect the index or the basic structure of the system.

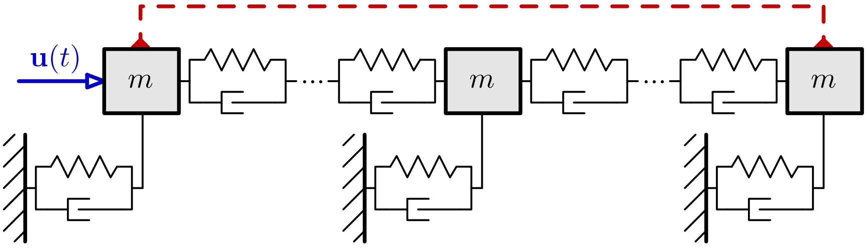


Figure 7.7: Constrained damped mass-spring system (adapted from [28, p. 106]). Each mass m is coupled with its neighbors and the environment by (linear) springs and dampers. The input $\mathbf{u}(t)$ (blue) directly controls the position of the first (counted from the left) mass. In contrast to [28], the holonomic constraint (red) between the first and last mass is removed.

The typical structure of holonomic[3] constrained linear multibody systems (as in Figure 7.7) looks like

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{x}}_1(t) \\ \dot{\mathbf{x}}_2(t) \\ \dot{\mathbf{x}}_3(t) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & -\mathbf{A}_{31}^{\mathrm{T}} \\ \mathbf{A}_{31} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \\ \mathbf{x}_3(t) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_2 \\ \mathbf{B}_3 \end{bmatrix} \mathbf{u}(t)\,, \tag{7.6}$$
$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \\ \mathbf{x}_3(t) \end{bmatrix}\,,$$

with $\mathbf{E}_{22}$ as the positive definite mass matrix, $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ as the stiffness and damping matrices respectively and $\mathbf{A}_{31}$ as the matrix encoding the algebraic constraints [9, p. 34]. The system state $\mathbf{x}(t)$ is composed of the displacement vector $\mathbf{x}_1(t)$, the velocity vector $\mathbf{x}_2(t)$ and the Lagrange multiplier $\mathbf{x}_3(t)$.

Note that in contrast to [9] the system stated in (7.6) is formulated with $\mathbf{B}_3 \neq \mathbf{0}$. In fact the special choice $\mathbf{B}_2 = \mathbf{0}$, $\mathbf{B}_3 = [0, \dots, 0, 1]^{\mathrm{T}}$ and $\mathbf{A}_{31} = [1, 0, \dots, 0]$ removes the constraint between the first and last mass. Simultaneously a direct coupling of $\mathbf{u}(t)$ and the position of the first mass is achieved. Keep in mind that the spectral projectors do not depend on the choice of $\mathbf{B}$ or $\mathbf{C}$ such that the analytic expressions from [9, p. 34] for $\mathbf{\Pi}_l^f$ and $\mathbf{\Pi}_r^f$ can be used without any modifications.

As in the previous section the actual benchmark model is generated by a MATLAB script created by Tatjana Stykel (at the Technische Universität Berlin in June 2006) which has been modified in order to incorporate the mentioned modifications of $\mathbf{B}_2$, $\mathbf{B}_3$ and $\mathbf{A}_{31}$. Note that models created according to [28] lead to a strictly proper transfer function. In order to obtain an improper benchmark model, the system output $\mathbf{y}(t)$ is chosen to be a combination of

[3] Usually the term "holonomic" is used to describe constraints between the system coordinates (i. e. the position of the masses). In the following this characterization is extended such that also constraints between system coordinates and the system input can be classified.

- the position of the third mass (counted from the left) causing a strictly proper contribution,
- the position of the first mass causing a proper contribution ($\mathcal{O}(\mathbf{P}(s)) \to 0$) and
- the velocity of the first mass causing an improper contribution ($\mathcal{O}(\mathbf{P}(s)) \to 1$).

Finally Figure 7.8 presents the frequency response of an accordingly generated FOM and its approximation obtained through Algorithm 6.1 (CURE aborted after two iterations).

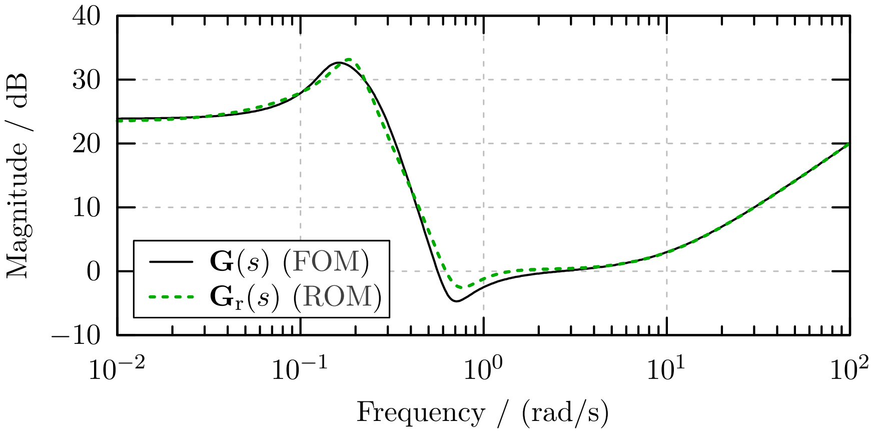


Figure 7.8: Frequency response of a mechanical index 3 DAE (illustrated in Figure 7.7) of dimension $n = 2001$ (corresponds to 1000 masses) and its approximation through a ROM of dimension $n_\mathrm{r} = 6$. Note that a rather low count of CURE iterations has been chosen in order to highlight the perfect matching of the improper subsystem.

As Figure 7.8 shows, the polynomial part $\mathbf{P}(s)$ is again matched exactly. This confirms that the proposed MOR-technique is suited for higher index and improper problems too.

7.4 Artificial High-Index System

In the previous sections physically based systems of different structure, index and properness have been investigated. Since in most technical applications the index of a DAE is limited to 3, an "artificial" benchmark model (i. e. without explicit physical interpretation) of high index is created and used to test the proposed reduction scheme in the following.

For this purpose a system with structure similar[4] to the Weierstraß canonical form is considered:

$$\begin{aligned}\begin{bmatrix}\mathbf{E}_f & \mathbf{0}\\ \mathbf{0} & \mathbf{E}_\infty\end{bmatrix}\begin{bmatrix}\dot{\mathbf{x}}_f(t)\\ \dot{\mathbf{x}}_\infty(t)\end{bmatrix} &= \begin{bmatrix}\mathbf{A}_f & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_\infty\end{bmatrix}\begin{bmatrix}\mathbf{x}_f(t)\\ \mathbf{x}_\infty(t)\end{bmatrix} + \begin{bmatrix}\mathbf{B}_f\\ \mathbf{B}_\infty\end{bmatrix}\mathbf{u}(t)\ ,\\ \mathbf{y}(t) &= \begin{bmatrix}\mathbf{C}_f & \mathbf{C}_\infty\end{bmatrix}\begin{bmatrix}\mathbf{x}_f(t)\\ \mathbf{x}_\infty(t)\end{bmatrix}\ .\end{aligned} \tag{7.7}$$

[4] In order to obtain a real-valued realization of the FOM, the matrix $\mathbf{A}_f$ is not in Jordan canonical form. Instead it is composed of an array of real-valued 2×2 blocks, each of them incorporating a complex conjugate eigenvalue pair.

Therein $[\mathbf{E}_f,\, \mathbf{A}_f,\, \mathbf{B}_f,\, \mathbf{C}_f]$ and $[\mathbf{E}_\infty,\, \mathbf{A}_\infty,\, \mathbf{B}_\infty,\, \mathbf{C}_\infty]$ denote realizations of the slow and fast subsystem of dimension n_f and n_∞ respectively. The spectral projectors are given through

$$\mathbf{\Pi}_l^f = \mathbf{\Pi}_r^f = \begin{bmatrix} \mathbf{I}_{n_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} , \qquad \mathbf{\Pi}_l^\infty = \mathbf{\Pi}_r^\infty = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix} . \tag{7.8}$$

In order to obtain a DAE-system of index ν (with $\nu \leq n_\infty$), one can use

$$\begin{aligned} E_{\infty,xy} &= \begin{cases} 1 & \text{for } (y = x+1) \wedge (x < \nu) \\ 0 & \text{otherwise} \end{cases} , & A_{\infty,xy} &= \begin{cases} 1 & \text{for } x = y \\ 0 & \text{otherwise} \end{cases} , \\ B_{\infty,x} &= \begin{cases} -1 & \text{for } x = \nu \\ 0 & \text{otherwise} \end{cases} , & C_{\infty,y} &= \begin{cases} \pi_{\nu-y} & \text{for } y \leq \nu \\ 0 & \text{otherwise} \end{cases} . \end{aligned} \tag{7.9}$$

to specify the entries of $\mathbf{E}_\infty$, $\mathbf{A}_\infty$, $\mathbf{B}_\infty$ and $\mathbf{C}_\infty$ in the x-th row and/or y-th column (with $x,\, y \in \{1,\, \dots,\, n_\infty\}$). Herein $\pi_0,\, \dots,\, \pi_{\nu-1} \in \mathbb{R}$ represent the coefficients of the resulting polynomial contribution $\mathbf{P}(s)$, i. e. $\mathbf{P}(s) = \pi_0 + \pi_1\, s + \pi_2\, s^2 + \dots + \pi_{\nu-1}\, s^{\nu-1}$.

In order to specify the slow subsystem, several complex conjugate eigenvalue pairs $\lambda_{w,1/2} = a_w \pm \imath\, b_w$ with $w \in \mathbb{N}^{>0}$ are considered, such that a corresponding 2×2 subsystem $\mathbf{\Sigma}_w = (\hat{\mathbf{E}}_{f,w},\, \hat{\mathbf{A}}_{f,w},\, \hat{\mathbf{B}}_{f,w},\, \hat{\mathbf{C}}_{f,w},\, \mathbf{0})$ can be formulated with

$$\hat{\mathbf{E}}_{f,w} = \mathbf{I}_2 \,, \quad \hat{\mathbf{A}}_{f,w} = \begin{bmatrix} a_w & b_w \\ -b_w & a_w \end{bmatrix} , \quad \hat{\mathbf{B}}_{f,w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} , \quad \hat{\mathbf{C}}_{f,w} = \begin{bmatrix} \frac{c_w}{2} \\ \frac{c_w}{2} \end{bmatrix} . \tag{7.10}$$

and $a_w,\, b_w,\, c_w \in \mathbb{R}$. Finally $\mathbf{E}_f$, $\mathbf{A}_f$, $\mathbf{B}_f$ and $\mathbf{C}_f$ have the block structure

$$\begin{aligned} \mathbf{E}_f &= \operatorname{diag}(\,\dots,\, \hat{\mathbf{E}}_{f,w},\, \dots\,) \,, & \mathbf{A}_f &= \operatorname{diag}(\,\dots,\, \hat{\mathbf{A}}_{f,w},\, \dots\,) \,, \\ \mathbf{B}_f &= [\,\dots,\, \hat{\mathbf{B}}_{f,w}^{\mathrm{T}},\, \dots]^{\mathrm{T}} \,, & \mathbf{C}_f &= [\,\dots,\, \hat{\mathbf{C}}_{f,w},\, \dots\,] \,. \end{aligned} \tag{7.11}$$

where each subsystem $\mathbf{\Sigma}_w$ is contained d_w times.

In the following the dimension and index of the investigated model are set to $n_f = 4000$, $n_\infty = 1000$ (thus $n = 5000$) and $\nu = 10$. Furthermore four different complex conjugate eigenvalue pairs (of multiplicity d_w) are considered. The actual choice of the parameters a_w, b_w, c_w, d_w and π_w is listed in Table 7.2.

Table 7.2: Parametrization of the transfer function $\mathbf{G}(s)$.

w	a_w	b_w	c_w	d_w	π_w
0					10^0
1	-10^{-3}	10^{-2}	$8 \cdot 10^{-2}$	$5 \cdot 10^2$	10^{-2}
2	-10^{-2}	10^{-1}	$1 \cdot 10^1$	$5 \cdot 10^2$	10^{-4}
3	-10^{-1}	10^0	$5 \cdot 10^1$	$5 \cdot 10^2$	10^{-5}
4	-10^0	10^1	$1 \cdot 10^3$	$5 \cdot 10^2$	10^{-5}
5					10^{-6}
6, 7, 8, 9					0

Note that the choice $\pi_5 \neq 0$ and $\pi_6 = \pi_7 = \pi_8 = \pi_9 = 0$ leads to $\mathcal{O}(\mathbf{P}(s)) = 5$. The transfer function of the FOM finally reads as

$$\mathbf{G}(s) = \sum_{w=1}^{4} d_w \, \frac{c_w \, (s - a_w)}{(s - a_w)^2 + b_w^2} + \sum_{w=0}^{5} \pi_w \, s^w \; . \tag{7.12}$$

Figure 7.9 shows the result of a reduction which has been aborted after 4 CURE iterations. Since $\mathcal{O}(\mathbf{P}(s)) = 5$ holds, the ROM is a DAE of index $\nu_\mathrm{r} = 6$[5] and order $n_\mathrm{r} = 14$[6].

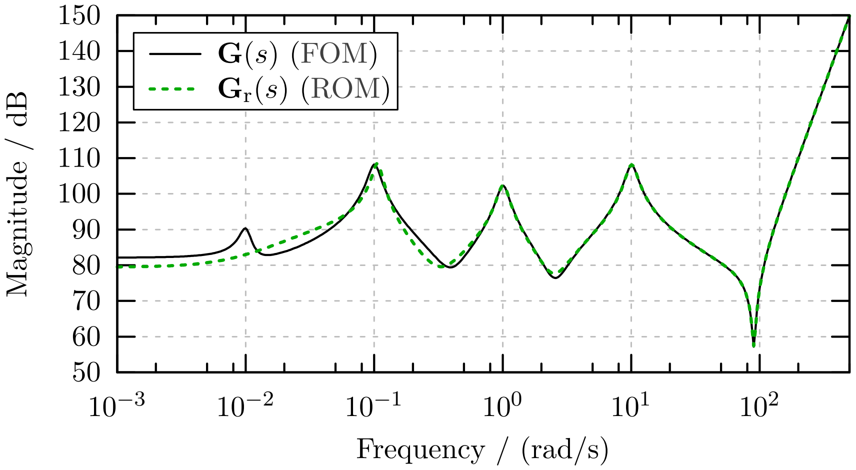


Figure 7.9: Frequency response of an index 10 DAE-system ($\mathcal{O}(\mathbf{P}(s)) = 5$) of dimension $n = 5000$ ($n_f = 4000$ and $n_\infty = 1000$) and its approximation through a ROM of order $n_\mathrm{r} = 14$ (4 CURE iterations).

As Figure 7.9 shows, the improper subsystem is matched even in the high index case. Since the proposed algorithm is not limited by the index of the FOM (at least in theory), one could have easily predicted this result. Anyway this example shows, that this basically holds in the numerical case too. Note that (as always in numerical mathematics) there are limits which originate form the calculation with finite machine precision. Thus there may be an upper limit for the index in practical applications.

[5] Keep in mind that $[\mathbf{E}_\mathrm{r}^\mathrm{im}, \mathbf{A}_\mathrm{r}^\mathrm{im}, \mathbf{B}_\mathrm{r}^\mathrm{im}, \mathbf{C}_\mathrm{r}^\mathrm{im}]$ obtained by Algorithm 5.2 is a minimal realization of $\mathbf{P}(s)$ (see Corollary 5.14). Thus ν_r has to be 6 in order to allow a polynomial contribution $\mathbf{P}_\mathrm{r}(s)$ of degree 5 (see (2.28)).

[6] $n_\mathrm{r} = 2 \cdot (\text{count of CURE iterations}) + \nu_\mathrm{r}$.

Chapter 8

Conclusions

The presented algorithm for the reduction of structured DAE-systems seems to be an elegant combination of Krylov subspace methods and SVD-based MOR. On the one hand BT is used to find a minimal realization of the polynomial part $\mathbf{P}(s)$, while on the other hand the strictly proper contributor $\mathbf{G}^{\mathrm{sp}}(s)$ is approximated by Krylov-based adaptive MOR through CUREd SPARK. Since the former affects the improper subsystem only, an efficient implementation is possible (i. e. without solving large-scale Lyapunov equations). As $\mathbf{P}(s)$ is matched exactly, the original problem of reducing an (improper) DAE-system changes to well-known ODE-MOR (i. e. to the approximation of a *rational* transfer function $\mathbf{G}^{\mathrm{sp}}(s)$).

Nevertheless even in the case of a strictly proper DAE-(sub)system, the descriptor matrix $\mathbf{E}$ is singular. Because of this one has to take special care during adaption of ODE-MOR-techniques to the DAE-case: although $\mathbf{G}(s)$ is represented by a rational function as in the ODE-case, the investigated MOR method might require regularity of $\mathbf{E}$. This is the case for the original derivation of the PORK algorithm given in [42], which makes use of $\mathbf{E}^{-1}$. Fortunately a modified proof has been found during this work, which is valid for singular $\mathbf{E}$ as well. Note that the restriction to strictly proper DAE-systems remains, since the $\mathcal{H}_2$ norm (and thus the concept of $\mathcal{H}_2$ pseudo-optimality) are only defined for this case. Thus the proposed partitioning into a slow and fast subsystem is inevitable for the reduction with PORK.

During the verification of the PORK algorithm for the DAE-case two necessary results of major importance have been derived. First, it has been shown, that the generalized Sylvester equation for the interpolation data

$$\mathbf{A}\,\mathbf{V} - \mathbf{E}\,\mathbf{V}\,\mathbf{S}_V = \mathbf{B}\,\mathbf{R} \tag{8.1}$$

has a unique solution $\mathbf{V}$, even in the case of singular $\mathbf{E}$. In fact the main condition for uniqueness of $\mathbf{V}$ is that the expansion points (encoded in $\mathbf{S}_V$) do not coincide with the generalized eigenvalues of the pair $(\mathbf{E},\,\mathbf{A})$. This allows the conclusion, that the representation of a tangential-input rational Krylov subspace in its original form is equivalent to the formulation as generalized Sylvester equation. Certainly the same holds for the dual case. Second, the $\mathcal{H}_2$ inner-product of the transfer functions of two asymptotically stable and strictly proper DAEs has been formulated by (projected) generalized Sylvester equations. This formulation plays an essential role in the proof of $\mathcal{H}_2$ pseudo-optimality in the PORK algorithm. As this is a very general result, it might be helpful in other contexts as well.

Beside the PORK algorithm also SPARK and its integration into the CURE-framework are applicable to the DAE-case. It is remarkable, that all three algorithms do not have to be modified and thus work “out of the box” with strictly proper DAEs. However the original derivation[1] of the SPARK algorithm from [30] has to be reformulated in order to comply with $\det(\mathbf{E}) = 0$.

As exemplified at the end of Section 7.1, the projection of $\mathbf{B}$ and $\mathbf{C}$ with the spectral projectors corresponding to the finite eigenvalues of $\lambda\,\mathbf{E} - \mathbf{A}$ is equivalent to finding the “underlying” ODE (as done in [11]). Both approaches represent an integration of the knowledge about the system structure into the MOR-technique. Although the formulation with spectral projectors used in this work seems to be more general, it might not be the most efficient way of reduction: even if analytic expressions for $\mathbf{\Pi}_l^f$ and $\mathbf{\Pi}_r^f$ are available, it might be possible to write down the contribution of the improper subsystem directly, at least in the case of semi-explicit index 1 DAEs (as shown in [18, p. B1021]). This way the SVD and subsequent truncation of zero improper HSVs can be avoided entirely.

Although the formulation with spectral projectors seems to be advantageous in many senses, it is essential that analytic expressions are available. Fortunately dedicated investigations have been made in the past for common system structures (several technical applications are collected in [9]). However if the spectral projectors are not known in advance a numerical expensive (and ill-conditioned) computation is required. This may destroy the benefits of the proposed algorithm in comparison to other MOR-techniques. Furthermore the matrices $\mathbf{\Pi}_l^f$ and $\mathbf{\Pi}_r^f$ are dense (in general), such that the reduction would be limited to small- or medium-sized problems.

With CUREd SPARK a powerful tool for adaptive, $\mathcal{H}_2$ pseudo-optimal reduction of SISO systems has been transferred to the DAE-world. In order to evaluate its performance, one has to compare the proposed algorithm to other MOR-techniques for DAEs such as IRKA (see [18]) or (entirely) SVD-based methods (see [9])). Unfortunately a comparison by means of numerical examples was not possible due to lack of time (although contained in the task-list of this thesis). At least a basic implementation of the proposed algorithm has been integrated into the sssMOR toolbox[2]. This issue should be tackled in future investigations.

Furthermore a detailed view on the matrix $[\mathbf{E}\,\mathbf{V},\, \mathbf{B}]$ and its column rank may be worthwhile, as it is a major requirement for the CURE framework (in the DAE-case as well as in the ODE-case). Especially universally valid statements could help to derive a MOR-scheme which satisfies this condition per construction.

[1] In fact solely the derivation of the analytic expressions for the gradient and Hessian (or more precisely: their abbreviated formulation) of the cost function has been revisited. The basic concept of SPARK, i. e. the maximization of $\|\mathbf{G}_\mathrm{r}\|_{\mathcal{H}_2}$, does not depend on $\mathbf{E}$ or its regularity at all.

[2] This feature of the toolbox has not been released for public use at the time of writing.

Appendix A

Proofs

Proof of Lemma 2.13

To prove part (i) the pencils $(\lambda\,\mathbf{X} - \mathbf{I}_u)$, $(\lambda\,\mathbf{Y} - \mathbf{I}_v)$ and $(\lambda\,\mathbf{Z} - \mathbf{I}_w)$ are transformed according to Lemma 2.5:

$$\mathbf{P}_x\,\mathbf{X}\,\mathbf{Q}_x = \tilde{\mathbf{X}}\,, \qquad \mathbf{P}_y\,\mathbf{Y}\,\mathbf{Q}_y = \tilde{\mathbf{Y}}\,, \qquad \mathbf{P}_z\,\mathbf{Z}\,\mathbf{Q}_z = \tilde{\mathbf{Z}}\,. \tag{A.1}$$

Stacking the transformation of $\lambda\,\mathbf{X} - \mathbf{I}_u$ and $\lambda\,\mathbf{Y} - \mathbf{I}_v$ into one equation leads to

$$\underbrace{\begin{bmatrix} \mathbf{P}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_y \end{bmatrix}}_{=:\mathbf{P}_{xy}} \underbrace{\begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y} \end{bmatrix}}_{\mathbf{Z}} \underbrace{\begin{bmatrix} \mathbf{Q}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_y \end{bmatrix}}_{=:\mathbf{Q}_{xy}} = \begin{bmatrix} \mathbf{I}_{u_f} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{v_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N}_y \end{bmatrix}\,. \tag{A.2}$$

Applying a (multiple) row-/column-swapping transformation with

$$\mathbf{T} := \begin{bmatrix} \mathbf{I}_{u_f} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{v_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{u_\infty} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{v_\infty} \end{bmatrix} \in \mathbb{R}^{w\times w}\,, \qquad \det(\mathbf{T}) = \pm 1 \tag{A.3}$$

delivers

$$\underbrace{(\mathbf{T}\,\mathbf{P}_{xy})}_{\mathbf{P}_z}\,\mathbf{Z}\,\underbrace{\left(\mathbf{Q}_{xy}\,\mathbf{T}^{\mathrm{T}}\right)}_{\mathbf{Q}_z} = \begin{bmatrix} \mathbf{I}_{u_f} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{v_f} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{N}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N}_y \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{w_f} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_z \end{bmatrix} = \tilde{\mathbf{Z}} \tag{A.4}$$

which is equivalent to the transformation of the matrix pencil $(\lambda\,\mathbf{Z} - \mathbf{I}_w)$ according to Lemma 2.5 with the transformation matrices $\mathbf{P}_z = \mathbf{T}\,\mathbf{P}_{xy}$ and $\mathbf{Q}_z = \mathbf{Q}_{xy}\,\mathbf{T}^{\mathrm{T}}$. Since $\mathbf{N}_z^{\eta_z}$ is blockdiagonal containing $\mathbf{N}_x^{\eta_z}$ and $\mathbf{N}_y^{\eta_z}$, it is zero, if both blocks are zero. Using this relation shows that $\eta_z = \max(\eta_x, \eta_y)$ which completes the proof of part (i).

In order to show part (ii) consider the three conditions in (2.6):

$$\begin{aligned}
\mathbf{Z}\,\mathbf{Z}^{\mathrm{D}} &= \operatorname{diag}(\mathbf{X},\,\mathbf{Y})\,\operatorname{diag}(\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}^{\mathrm{D}}) = \operatorname{diag}(\mathbf{X}\,\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}\,\mathbf{Y}^{\mathrm{D}})\\
&= \operatorname{diag}(\mathbf{X}^{\mathrm{D}}\,\mathbf{X},\,\mathbf{Y}^{\mathrm{D}}\,\mathbf{Y}) = \operatorname{diag}(\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}^{\mathrm{D}})\,\operatorname{diag}(\mathbf{X},\,\mathbf{Y}) = \mathbf{Z}^{\mathrm{D}}\,\mathbf{Z}\,,\\[1ex]
\mathbf{Z}^{\mathrm{D}}\,\mathbf{Z}\,\mathbf{Z}^{\mathrm{D}} &= \operatorname{diag}(\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}^{\mathrm{D}})\,\operatorname{diag}(\mathbf{X},\,\mathbf{Y})\,\operatorname{diag}(\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}^{\mathrm{D}})\\
&= \operatorname{diag}(\mathbf{X}^{\mathrm{D}}\,\mathbf{X}\,\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}^{\mathrm{D}}\,\mathbf{Y}\,\mathbf{Y}^{\mathrm{D}}) = \operatorname{diag}(\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}^{\mathrm{D}}) = \mathbf{Z}^{\mathrm{D}}\,,\\[1ex]
\mathbf{Z}^{\mathrm{D}}\,\mathbf{Z}^{\eta_z+1} &= \operatorname{diag}(\mathbf{X}^{\mathrm{D}},\,\mathbf{Y}^{\mathrm{D}})\,\operatorname{diag}(\mathbf{X}^{\eta_z+1},\,\mathbf{Y}^{\eta_z+1}) = \operatorname{diag}(\mathbf{X}^{\mathrm{D}}\,\mathbf{X}^{\eta_z+1},\,\mathbf{Y}^{\mathrm{D}}\,\mathbf{Y}^{\eta_z+1})\\
&\overset{\text{(i)}}{=} \operatorname{diag}(\mathbf{X}^{\eta_x}\,\mathbf{X}^{\eta_z-\eta_x},\,\mathbf{Y}^{\eta_y}\,\mathbf{Y}^{\eta_z-\eta_y}) = \operatorname{diag}(\mathbf{X}^{\eta_z},\,\mathbf{Y}^{\eta_z}) = \mathbf{Z}^{\eta_z}\,.
\end{aligned} \tag{A.5}$$

In the third equation result (i) was used, to ensure, that $\eta_z \geq \eta_x$ and $\eta_z \geq \eta_y$.

An alternative proof of part (ii) as well as a relation similar to part (i) is included in [13, p. 2772] which uses results from [14, p. 632]. ■

Proof of Key Theorem 4.17

As shown in Corollary 4.10[1] there exists a *unique* $\mathbf{X}$ as the solution of

$$\mathbf{A}\,\mathbf{X}\,\mathbf{E}_{\mathrm{M}}^{\mathfrak{l}*} + \mathbf{E}\,\mathbf{X}\,\mathbf{A}_{\mathrm{M}}^{\mathfrak{l}*} + \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{B}_{\mathrm{M}}^{\mathfrak{l}*} = \mathbf{0}\,, \tag{A.6}$$

such that $\langle \mathbf{G},\,\mathbf{G}_{\mathrm{M}}\rangle_{\mathcal{H}_2} = \operatorname{tr}(\mathbf{C}\,\mathbf{X}\,\mathbf{C}_{\mathrm{M}}^{\mathfrak{l}*})$ holds. An appropriate partitioning of $\mathbf{X}$

$$\mathbf{X} = [\mathbf{X}_1,\,\dots,\,\mathbf{X}_i,\,\dots,\,\mathbf{X}_s]\,, \tag{A.7a}$$

$$\mathbf{X}_i = [\mathbf{X}_{i1},\,\dots,\,\mathbf{X}_{ij},\,\dots,\,\mathbf{X}_{ir_i}]\,, \qquad \forall\, i = 1,\,\dots,\,s \tag{A.7b}$$

$$\mathbf{X}_{ij} = \left[\mathbf{x}_{ij1},\,\dots,\,\mathbf{x}_{ijk},\,\dots,\,\mathbf{x}_{ijq_{ij}}\right]\,, \qquad \forall\, j = 1,\,\dots,\,r_i \tag{A.7c}$$

with $\mathbf{x}_{ijk} \in \mathbb{C}^{n\times 1}$ leads together with $\mathbf{E}_{\mathrm{M}}^{\mathfrak{l}} = \mathbf{I}_q$ to

$$\mathbf{A}\,\mathbf{X}_{ij} + \mathbf{E}\,\mathbf{X}_{ij}\mathbf{A}_{\mathrm{M}ij}^{\mathfrak{l}*} + \mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{B}_{\mathrm{M}ij}^{\mathfrak{l}*} = \mathbf{0}\,. \tag{A.8}$$

A further decomposition delivers

$$\begin{aligned}
&\mathbf{A}\left[\mathbf{x}_{ij1},\,\dots,\,\mathbf{x}_{ijk},\,\dots,\,\mathbf{x}_{ijq_{ij}}\right] + \mathbf{E}\left[\mathbf{x}_{ij1},\,\dots,\,\mathbf{x}_{ijk},\,\dots,\,\mathbf{x}_{ijq_{ij}}\right]\begin{bmatrix}\overline{\lambda}_{\mathrm{M}i} & -1 & & \\ & \ddots & \ddots & \\ & & \ddots & -1\\ & & & \overline{\lambda}_{\mathrm{M}i}\end{bmatrix} +\\
&\quad + \mathbf{\Pi}_l^f\,\mathbf{B}\left[\mathbf{b}_{\mathrm{M}ij1}^{\mathfrak{l}*},\,\dots,\,\mathbf{b}_{\mathrm{M}ijk}^{\mathfrak{l}*},\,\dots,\,\mathbf{b}_{\mathrm{M}ijq_{ij}}^{\mathfrak{l}*}\right] = \mathbf{0}\,,
\end{aligned} \tag{A.9}$$

and therefore

$$\mathbf{x}_{ijk} = \begin{cases} -\left(\mathbf{A} + \overline{\lambda}_{\mathrm{M}i}\,\mathbf{E}\right)^{-1}\mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{b}_{\mathrm{M}ijk}^{\mathfrak{l}*}\,, & \text{for } k = 1\,,\\ \left(\mathbf{A} + \overline{\lambda}_{\mathrm{M}i}\,\mathbf{E}\right)^{-1}\mathbf{E}\,\mathbf{x}_{ij(k-1)} - \left(\mathbf{A} + \overline{\lambda}_{\mathrm{M}i}\,\mathbf{E}\right)^{-1}\mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{b}_{\mathrm{M}ijk}^{\mathfrak{l}*}\,, & \text{for } 1 < k \leq q_{ij}\,. \end{cases}$$

[1] Note that $\mathbf{G}_{\mathrm{M}}(s)$ is of ODE-type since $\det(\mathbf{E}_{\mathrm{M}}^{\mathfrak{l}}) = \det(\mathbf{I}_q) \neq 0$.

(A.10)

Note that since both transfer functions, $\mathbf{G}(s)$ and $\mathbf{G}_{\mathrm{M}}(s)$, belong to asymptotically stable systems, the matrix $\left(\mathbf{A}+\overline{\lambda}_{\mathrm{M}i}\,\mathbf{E}\right)$ is guaranteed to be invertible.

The recursion can be reformulated as

$$\mathbf{x}_{ijk} = -\sum_{\xi=1}^{k} \left[\left(\mathbf{A}+\overline{\lambda}_{\mathrm{M}i}\,\mathbf{E}\right)^{-1}\,\mathbf{E}\right]^{k-\xi} \left(\mathbf{A}+\overline{\lambda}_{\mathrm{M}i}\,\mathbf{E}\right)^{-1}\,\mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}\,. \tag{A.11}$$

The $\mathcal{H}_2$ inner-product finally reads as

$$\begin{aligned}
\langle \mathbf{G},\,\mathbf{G}_{\mathrm{M}}\rangle_{\mathcal{H}_2} &= \mathrm{tr}\left(\mathbf{C}\,\mathbf{X}\,\mathbf{C}_{\mathrm{M}}^{\mathfrak{l}*}\right) = \mathrm{tr}\left(\mathbf{C}\sum_{i=1}^{s}\sum_{j=1}^{r_i}\sum_{k=1}^{q_{ij}}\mathbf{x}_{ijk}\,\mathbf{c}_{\mathrm{M}ijk}^{\mathfrak{l}*}\right)\\
&= -\,\mathrm{tr}\left(\sum_{i=1}^{s}\sum_{j=1}^{r_i}\sum_{k=1}^{q_{ij}}\sum_{\xi=1}^{k}\mathbf{C}\left[(\mathbf{A}+\overline{\lambda}_{\mathrm{M}i}\,\mathbf{E})^{-1}\,\mathbf{E}\right]^{k-\xi}(\mathbf{A}+\overline{\lambda}_{\mathrm{M}i}\,\mathbf{E})^{-1}\mathbf{\Pi}_l^f\,\mathbf{B}\,\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}\,\mathbf{c}_{\mathrm{M}ijk}^{\mathfrak{l}*}\right)\\
&\overset{(2.35)}{=} -\,\mathrm{tr}\left(\sum_{i=1}^{s}\sum_{j=1}^{r_i}\sum_{k=1}^{q_{ij}}\sum_{\xi=1}^{k}\mathbf{M}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})\,\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}\,\mathbf{c}_{\mathrm{M}ijk}^{\mathfrak{l}*}\right)\\
&\overset{(\text{Lemma B.1})}{=} -\,\mathrm{tr}\left(\sum_{i=1}^{s}\sum_{j=1}^{r_i}\sum_{k=1}^{q_{ij}}\sum_{\xi=1}^{k}\mathbf{c}_{\mathrm{M}ijk}^{\mathfrak{l}*}\,\mathbf{M}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})\,\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}\right)\\
&= -\sum_{i=1}^{s}\sum_{j=1}^{r_i}\sum_{k=1}^{q_{ij}}\sum_{\xi=1}^{k}\mathbf{c}_{\mathrm{M}ijk}^{\mathfrak{l}*}\,\mathbf{M}^{(k-\xi)}(-\overline{\lambda}_{\mathrm{M}i})\,\mathbf{b}_{\mathrm{M}ij\xi}^{\mathfrak{l}*}
\end{aligned} \tag{A.12}$$

which completes the proof. ■

Proof of Theorem 4.22

Since $\mathbf{G}_{\mathrm{F}}(s)$ can be described by the interpolation matrices of the primitive basis $\mathbf{V}^{\mathrm{P}}$ (see Remark 4.21), one can make use of the special structure of $\mathbf{S}_V^{\mathrm{P}}$ and $\mathbf{R}^{\mathrm{P}}$:

$$\begin{aligned}
\mathbf{S}_V^{\mathrm{P}} &= \mathrm{diag}\left(\mathbf{S}_{V,1}^{\mathrm{P}},\,\dots,\,\mathbf{S}_{V,i}^{\mathrm{P}},\,\dots,\,\mathbf{S}_{V,s}^{\mathrm{P}}\right)\,, & \mathbf{R}^{\mathrm{P}} &= \left[\mathbf{R}_1^{\mathrm{P}},\,\dots,\,\mathbf{R}_i^{\mathrm{P}},\,\dots,\,\mathbf{R}_s^{\mathrm{P}}\right]\,,\\
\mathbf{S}_{V,i}^{\mathrm{P}} &= \mathrm{diag}\left(\mathbf{S}_{V,i1}^{\mathrm{P}},\,\dots,\,\mathbf{S}_{V,ij}^{\mathrm{P}},\,\dots,\,\mathbf{S}_{V,ir_i}^{\mathrm{P}}\right)\,, & \mathbf{R}_i^{\mathrm{P}} &= \left[\mathbf{R}_{i1}^{\mathrm{P}},\,\dots,\,\mathbf{R}_{ij}^{\mathrm{P}},\,\dots,\,\mathbf{R}_{ir_i}^{\mathrm{P}}\right]\,,\\
\mathbf{S}_{V,ij}^{\mathrm{P}} &= \begin{bmatrix} s_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & s_i\end{bmatrix} \in \mathbb{C}^{q_{ij}\times q_{ij}}\,, & \mathbf{R}_{ij}^{\mathrm{P}} &= [\mathbf{r}_{ij},\,\mathbf{0},\,\dots,\,\mathbf{0}] \in \mathbb{C}^{m\times q_{ij}}\,.
\end{aligned} \tag{A.13}$$

In order to prove tangential interpolation at the particular expansion point s_i, the structure of the matrix

$$\mathbf{K} := \mathbf{S}_V^{\mathrm{P}} + \mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}} - s_i\,\mathbf{I}_q \tag{A.14}$$

is analyzed. For this purpose one can find, that the product $\mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}}$ has for any $\mathbf{F}^{\mathrm{P}} \in \mathbb{C}^{q\times m}$ the form

$$\mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}} = \left[\underset{\underset{\mathbf{F}^{\mathrm{P}}\,\mathbf{r}_{11}}{\uparrow}}{\begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix}} \;\cdots\; \underset{\underset{\mathbf{F}^{\mathrm{P}}\,\mathbf{r}_{sr_s}}{\uparrow}}{\begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix}}\right] \in \mathbb{C}^{q\times q}\,, \tag{A.15}$$

wherein $*$ represents an arbitrary scalar value. Similarly, the structure of $\mathbf{S}_V^{\mathrm{P}} - s_i\,\mathbf{I}_q$ can be visualized:

$$\mathbf{S}_V^{\mathrm{P}} - s_i\,\mathbf{I}_q = \left[\begin{array}{ccc} \begin{pmatrix} * & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & * \end{pmatrix} & & \\ & \ddots & \\ & & \begin{pmatrix} * & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & * \end{pmatrix} \\ \uparrow & & \uparrow \\ \mathbf{S}_{V,11}^{\mathrm{P}} - s_i\,\mathbf{I}_{q_{11}} & & \mathbf{S}_{V,sr_s}^{\mathrm{P}} - s_i\,\mathbf{I}_{q_{sr_s}} \end{array}\right] \in \mathbb{C}^{q\times q}\,. \tag{A.16}$$

Finally the overall structure of $\mathbf{K}$ reads as

$$\mathbf{K} = \left[\begin{array}{cccc} \begin{pmatrix} * & 1 & & \\ * & * & \ddots & \\ \vdots & & \ddots & 1 \\ * & & & * \end{pmatrix} & \begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix} & \cdots & \begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix} \\ \begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix} \\ \begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix} & \cdots & \begin{pmatrix} * & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0\end{pmatrix} & \begin{pmatrix} * & 1 & & \\ * & * & \ddots & \\ \vdots & & \ddots & 1 \\ * & & & * \end{pmatrix} \end{array}\right] \in \mathbb{C}^{q\times q}\,, \tag{A.17}$$

where each column of blocks corresponds to a special choice of expansion point and tangential direction.

Let $\mathbf{e}_\gamma$ denote the γ-th unit vector in $\mathbb{R}^q$, i. e.

$$\mathbf{e}_\gamma = [e_{\gamma,1},\, \dots,\, e_{\gamma,u},\, \dots,\, e_{\gamma,q}]^{\mathrm{T}} \in \mathbb{R}^{q\times 1} \quad \text{with} \quad e_{\gamma,u} = \begin{cases} 1 & \text{for } u = \gamma \\ 0 & \text{for } u \neq \gamma \end{cases}\,. \tag{A.18}$$

Multiplying $\mathbf{K}$ from the right with $\mathbf{e}_\gamma$ allows to extract the γ-th column $\mathbf{k}_\gamma \in \mathbb{C}^{q\times 1}$. In order to get the k-th column of the blocks related to the j-th tangential direction of the i-th expansion point, one can use the mapping

$$\gamma(i,\, j,\, k) = \sum_{u=1}^{i-1} \sum_{v=1}^{r_u} q_{uv} + \sum_{w=1}^{j-1} q_{iw} + k \; . \tag{A.19}$$

Now put the focus on the columns of $\mathbf{K}$ corresponding to the i-th expansion point (s_i) and the j-th tangential direction ($\mathbf{r}_{ij}$). Since $\mathbf{S}_V^{\mathrm{P}}$ is in Jordan canonical form, the diagonal elements of $\mathbf{S}_V^{\mathrm{P}} - s_i\,\mathbf{I}_q$ related to the i-th expansion point vanish and it follows that

$$\mathbf{k}_\gamma := \mathbf{K}\,\mathbf{e}_\gamma = \begin{cases} \mathbf{F}^{\mathrm{P}}\,\mathbf{r}_{ij}\;, & \text{for } \gamma = \gamma(i,\, j,\, k=1) \\ \mathbf{e}_{\gamma-1}\;, & \text{for } \gamma = \gamma(i,\, j,\, k>1) \end{cases} \; . \tag{A.20}$$

Inserting (A.14) back into (A.20) leads to a series of LSEs

$$\begin{aligned} &\left(\mathbf{S}_V^{\mathrm{P}} + \mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}} - s_i\,\mathbf{I}_q\right)\mathbf{e}_{\gamma(i,\,j,\,1)} = \mathbf{F}^{\mathrm{P}}\,\mathbf{r}_{ij}\;, \\ &\left(\mathbf{S}_V^{\mathrm{P}} + \mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}} - s_i\,\mathbf{I}_q\right)\mathbf{e}_{\gamma(i,\,j,\,w)} = \mathbf{e}_{\gamma(i,\,j,\,w-1)}\;, \qquad \forall\, w = 2,\, \dots,\, q_{ij} \end{aligned} \tag{A.21}$$

which have unique solutions $\mathbf{e}_\gamma$, since $\lambda(\mathbf{S}_V) \cap \lambda(\mathbf{E}_{\mathrm{F}} = \mathbf{I}_q,\, \mathbf{A}_{\mathrm{F}}) = \emptyset$ and therefore $\lambda(\mathbf{S}_V^{\mathrm{P}}) \cap \lambda(\mathbf{E}_{\mathrm{F}}^{\mathrm{P}} = \mathbf{I}_q,\, \mathbf{A}_{\mathrm{F}}^{\mathrm{P}}) = \emptyset$[2] holds.

Recursively solving (A.21) leads to

$$\mathbf{e}_{\gamma(i,\,j,\,k)} = \left(\mathbf{S}_V^{\mathrm{P}} + \mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}} - s_i\,\mathbf{I}_q\right)^{-k}\mathbf{F}^{\mathrm{P}}\,\mathbf{r}_{ij}\;, \qquad \forall\, k = 1,\, \dots,\, q_{ij}\; . \tag{A.22}$$

Using the definition of the moments of a transfer function in (2.34) together with (4.41) delivers

$$\begin{aligned} \left(\frac{\mathrm{d}^\mu \mathbf{G}_{\mathrm{F}}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} &= -(\mu!)\,\mathbf{C}_{\mathrm{F}}\,\left[(\mathbf{A}_{\mathrm{F}} - s_i\,\mathbf{E}_{\mathrm{F}})^{-1}\,\mathbf{E}_{\mathrm{F}}\right]^\mu\,(\mathbf{A}_{\mathrm{F}} - s_i\,\mathbf{E}_{\mathrm{F}})^{-1}\,\mathbf{B}_{\mathrm{F}}\,\mathbf{r}_{ij} \\ &= -(\mu!)\,\mathbf{C}_{\mathrm{F}}^{\mathrm{P}}\,\left(\mathbf{A}_{\mathrm{F}}^{\mathrm{P}} - s_i\,\mathbf{I}_q\right)^{-(\mu+1)}\,\mathbf{B}_{\mathrm{F}}^{\mathrm{P}}\,\mathbf{r}_{ij} \\ &= -(\mu!)\,\mathbf{C}\,\mathbf{V}^{\mathrm{P}}\,\left(\mathbf{S}_V^{\mathrm{P}} + \mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}} - s_i\,\mathbf{I}_q\right)^{-(\mu+1)}\,\mathbf{F}^{\mathrm{P}}\,\mathbf{r}_{ij}\; . \end{aligned} \tag{A.23}$$

Inserting (A.22) leads to

$$\left(\frac{\mathrm{d}^\mu \mathbf{G}_{\mathrm{F}}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} = -(\mu!)\,\mathbf{C}\,\mathbf{V}^{\mathrm{P}}\,\mathbf{e}_{\gamma(i,\,j,\,\mu+1)}\;, \qquad \forall\, \mu = 0,\, \dots,\, q_{ij} - 1\; . \tag{A.24}$$

Note that $\mathbf{V}^{\mathrm{P}}\,\mathbf{e}_{\gamma(i,\,j,\,\mu+1)}$ denotes the $\gamma(i,\, j,\, \mu+1)$-th column of the primitive base $\mathbf{V}^{\mathrm{P}}$ of $\mathcal{K}_{\mathrm{ti}}$. Therefore

$$\begin{aligned} \left(\frac{\mathrm{d}^\mu \mathbf{G}_{\mathrm{F}}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} &= -(\mu!)\,\mathbf{C}\,\left[(\mathbf{A} - s_i\,\mathbf{E})^{-1}\mathbf{E}\right]^\mu\,(\mathbf{A} - s_i\,\mathbf{E})^{-1}\mathbf{B}\,\mathbf{r}_{ij} \\ &= \left(\frac{\mathrm{d}^\mu \mathbf{G}(s)}{\mathrm{d}s^\mu}\right)\bigg|_{s=s_i} \cdot \mathbf{r}_{ij} \end{aligned} \tag{A.25}$$

is fulfilled for all $\mu = 0,\, \dots,\, q_{ij}$. Since the above considerations hold for all expansion points s_i (with $i = 1,\, \dots,\, s$) and all tangential directions $\mathbf{r}_{ij}$ (with $j = 1,\, \dots,\, r_i$), the proof is complete. ■

[2] Because $\mathbf{S}_V$ and $\mathbf{S}_V^{\mathrm{P}}$ are similar, they share the same set of eigenvalues. The same holds for $\mathbf{A}_{\mathrm{F}} = \mathbf{S}_V + \mathbf{F}\,\mathbf{R}$ and $\mathbf{A}_{\mathrm{F}}^{\mathrm{P}} = \mathbf{S}_V^{\mathrm{P}} + \mathbf{F}^{\mathrm{P}}\,\mathbf{R}^{\mathrm{P}}$ with $\mathbf{A}_{\mathrm{F}} = \mathbf{T}^{-1}\,\mathbf{A}_{\mathrm{F}}^{\mathrm{P}}\,\mathbf{T}$.

Appendix B

Secondary Results

Lemma B.1. *Let* $\mathbf{X} \in \mathbb{C}^{v\times u}$ *and* $\mathbf{Y} \in \mathbb{C}^{u\times v}$. *Then* $\operatorname{tr}(\mathbf{X}\,\mathbf{Y}) = \operatorname{tr}(\mathbf{Y}\,\mathbf{X})$ *holds.*

Proof. The explicit computation of the trace

$$\begin{aligned}\operatorname{tr}(\mathbf{X}\,\mathbf{Y}) &= \sum_{x=1}^{v} (\mathbf{X}\,\mathbf{Y})_{xx} = \sum_{x=1}^{v}\sum_{y=1}^{u} X_{xy}\,Y_{yx} = \sum_{y=1}^{u}\sum_{x=1}^{v} Y_{yx}\,X_{xy} \\ &= \sum_{y=1}^{u} (\mathbf{Y}\,\mathbf{X})_{yy} = \operatorname{tr}(\mathbf{Y}\,\mathbf{X})\end{aligned} \tag{B.1}$$

verifies the claim. ∎

Lemma B.2. *Let*

- $(\mathbf{E},\,\mathbf{A},\,\mathbf{B},\,\mathbf{C},\,\mathbf{x}_0)$ *be a* DAE*-system according to Definition 2.14 with regular matrix pencil* $\lambda\,\mathbf{E} - \mathbf{A}$ *and transfer function* $\mathbf{G}(s)$*,*
- $[\tilde{\mathbf{E}},\,\tilde{\mathbf{A}},\,\tilde{\mathbf{B}},\,\tilde{\mathbf{C}}]$ *be a realization of* $\mathbf{G}(s)$ *in Weierstraß canonical form where* $\mathbf{P}$ *and* $\mathbf{Q}$ *are the transformation matrices as given in Lemma 2.5,*
- $\mathbf{\Pi}_l^f$, $\mathbf{\Pi}_l^\infty$, $\mathbf{\Pi}_r^f$, $\mathbf{\Pi}_r^\infty$ *be the spectral projectors according to Definition 2.8 and*
- s *be an arbitrary complex-valued scalar with* $s \notin \lambda(\mathbf{E},\,\mathbf{A})$.

Then the equalities

$$\mathbf{\Pi}_l^f\,\mathbf{A} = \mathbf{A}\,\mathbf{\Pi}_r^f\,, \quad \mathbf{\Pi}_l^\infty\,\mathbf{A} = \mathbf{A}\,\mathbf{\Pi}_r^\infty\,, \quad \mathbf{\Pi}_l^f\,\mathbf{E} = \mathbf{E}\,\mathbf{\Pi}_r^f\,, \quad \mathbf{\Pi}_l^\infty\,\mathbf{E} = \mathbf{E}\,\mathbf{\Pi}_r^\infty\,, \tag{B.2}$$

$$\mathbf{\Pi}_r^f\,(s\,\mathbf{E} - \mathbf{A})^{-1} = (s\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{\Pi}_l^f \quad \textit{and} \tag{B.3}$$

$$\mathbf{\Pi}_r^\infty\,(s\,\mathbf{E} - \mathbf{A})^{-1} = (s\,\mathbf{E} - \mathbf{A})^{-1}\,\mathbf{\Pi}_l^\infty \tag{B.4}$$

hold.

Proof. To prove the relations in (B.2) consider

$$\begin{aligned}\mathbf{\Pi}_l^f\,\mathbf{A} &= \mathbf{P}^{-1}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\underbrace{\mathbf{P}\,\mathbf{A}\,\mathbf{Q}}_{\tilde{\mathbf{A}}}\,\mathbf{Q}^{-1} = \mathbf{P}^{-1}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{J} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_\infty}\end{bmatrix}\mathbf{Q}^{-1} \\ &= \mathbf{P}^{-1}\begin{bmatrix}\mathbf{J} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1}\end{aligned} \tag{B.5}$$

and

$$\begin{aligned}\mathbf{A}\,\mathbf{\Pi}_r^f &= \mathbf{P}^{-1}\,\underbrace{\mathbf{P}\,\mathbf{A}\,\mathbf{Q}}_{\tilde{\mathbf{A}}}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1} = \mathbf{P}^{-1}\begin{bmatrix}\mathbf{J} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_\infty}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1}\\ &= \mathbf{P}^{-1}\begin{bmatrix}\mathbf{J} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{Q}^{-1}\end{aligned} \tag{B.6}$$

which shows, that $\mathbf{\Pi}_l^f\,\mathbf{A} = \mathbf{A}\,\mathbf{\Pi}_r^f$ holds. The remaining equalities of (B.2) can be shown in the same manner.

In order to verify (B.3) a similar method is used:

$$\begin{aligned}&\mathbf{\Pi}_r^f\,(s\,\mathbf{E}-\mathbf{A})^{-1} = \mathbf{\Pi}_r^f\,\mathbf{Q}\,(s\,\mathbf{P}\,\mathbf{E}\,\mathbf{Q}-\mathbf{P}\,\mathbf{A}\,\mathbf{Q})^{-1}\,\mathbf{P}\\ &\quad= \mathbf{Q}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\left(s\,\mathbf{I}_{n_f}-\mathbf{J}\right)^{-1} & \mathbf{0}\\ \mathbf{0} & (s\,\mathbf{N}-\mathbf{I}_{n_\infty})^{-1}\end{bmatrix}\mathbf{P} = \mathbf{Q}\begin{bmatrix}\left(s\,\mathbf{I}_{n_f}-\mathbf{J}\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}\end{aligned} \tag{B.7}$$

with

$$\begin{aligned}&(s\,\mathbf{E}-\mathbf{A})^{-1}\,\mathbf{\Pi}_l^f = \mathbf{Q}\,(s\,\mathbf{P}\,\mathbf{E}\,\mathbf{Q}-\mathbf{P}\,\mathbf{A}\,\mathbf{Q})^{-1}\,\mathbf{P}\,\mathbf{\Pi}_l^f\\ &\quad= \mathbf{Q}\begin{bmatrix}\left(s\,\mathbf{I}_{n_f}-\mathbf{J}\right)^{-1} & \mathbf{0}\\ \mathbf{0} & (s\,\mathbf{N}-\mathbf{I}_{n_\infty})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P} = \mathbf{Q}\begin{bmatrix}\left(s\,\mathbf{I}_{n_f}-\mathbf{J}\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{P}\end{aligned} \tag{B.8}$$

such, that $\mathbf{\Pi}_r^f\,(s\,\mathbf{E}-\mathbf{A})^{-1} = (s\,\mathbf{E}-\mathbf{A})^{-1}\,\mathbf{\Pi}_l^f$ is shown. The proof of (B.4) is analogous. ■

Corollary B.3. *Let all conditions of Lemma B.2, Theorem 3.14 and Corollary 3.15 hold. If additionally* $\mathbf{B} = \mathbf{\Pi}_l^f\,\mathbf{B}$ *is fulfilled, then the equalities*

$$\mathbf{\Pi}_r^f\,\mathbf{V}^{\mathrm{P}} = \mathbf{V}^{\mathrm{P}} \quad \textit{and} \quad \mathbf{\Pi}_r^f\,\mathbf{V} = \mathbf{V} \tag{B.9}$$

hold.

Proof. Consider the definition of the primitive basis $\mathbf{V}^{\mathrm{P}}$ from Section 3.2:

$$\mathbf{V}^{\mathrm{P}} = [\,\dots,\,\mathbf{v}_{ijk}^{\mathrm{P}},\,\dots\,]\,, \qquad \text{with } \mathbf{v}_{ijk}^{\mathrm{P}} = \left[(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\mathbf{E}\right]^{k-1}(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\mathbf{B}\,\mathbf{r}_{ij}\,. \tag{B.10}$$

Using the results from Lemma B.2 allows to “pass” the spectral projector through until it hits **B**:

$$\begin{aligned}&\mathbf{\Pi}_r^f\,\mathbf{V}^{\mathrm{P}} = [\,\dots,\,\mathbf{\Pi}_r^f\,\mathbf{v}_{ijk}^{\mathrm{P}},\,\dots\,]\,,\\ &\mathbf{\Pi}_r^f\,\mathbf{v}_{ijk}^{\mathrm{P}} = \left[(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\mathbf{E}\right]^{k-1}(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\underbrace{\mathbf{\Pi}_l^f\,\mathbf{B}}_{\mathbf{B}}\,\mathbf{r}_{ij} = \mathbf{v}_{ijk}^{\mathrm{P}}\,.\end{aligned} \tag{B.11}$$

Since $\mathbf{B} = \mathbf{\Pi}_l^f\,\mathbf{B}$ holds, the equality $\mathbf{\Pi}_r^f\,\mathbf{V}^{\mathrm{P}} = \mathbf{V}^{\mathrm{P}}$ is shown. Furthermore, the transformation $\mathbf{V} = \mathbf{V}^{\mathrm{P}}\,\mathbf{T}$ from Corollary 3.15 allows to verify

$$\mathbf{\Pi}_r^f\,\mathbf{V} = \mathbf{\Pi}_r^f\,\mathbf{V}^{\mathrm{P}}\,\mathbf{T} = \mathbf{V}^{\mathrm{P}}\,\mathbf{T} = \mathbf{V}\,, \tag{B.12}$$

which completes the proof. ■

Lemma B.4. *Let* $(\mathbf{E},\ \mathbf{A},\ \mathbf{B},\ \mathbf{C},\ \mathbf{x}_0)$ *be an asymptotically stable DAE-system with regular matrix pencil* $\lambda\,\mathbf{E} - \mathbf{A}$*. Then the matrix* $\mathbf{A}$ *is regular, i. e.* $\det(\mathbf{A}) \neq 0$*.*

Proof. Consider the transformation of $\mathbf{A}$ into Weierstraß canonical form with the regular matrices $\mathbf{P}$ and $\mathbf{Q}$ according to Lemma 2.5:

$$\tilde{\mathbf{A}} = \mathbf{P}\,\mathbf{A}\,\mathbf{Q} = \begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix} , \quad \text{with} \quad \det(\mathbf{P}),\, \det\mathbf{Q} \neq 0\,. \tag{B.13}$$

Since $\mathbf{J}$ is in Jordan canonical form and $\lambda(\mathbf{J}) = \lambda_f(\mathbf{E},\ \mathbf{A})$ (see Lemma 2.5) and $0 \notin \lambda_f(\mathbf{E},\ \mathbf{A})$ (asymptotically stable), all diagonal elements of the upper triangular matrix $\tilde{\mathbf{A}}$ are nonzero. Thus $\det(\mathbf{A}) = \det(\mathbf{P}^{-1}\,\tilde{\mathbf{A}}\,\mathbf{Q}^{-1}) = \det(\mathbf{P}^{-1})\,\det(\tilde{\mathbf{A}})\,\det(\mathbf{Q}^{-1}) \neq 0$ holds. ■

Corollary B.5. *Let all conditions of Lemma B.2 and Lemma B.4 hold. Moreover let* ν *denote the index of the matrix pencil* $\lambda\,\mathbf{E} - \mathbf{A}$ *according to Definition 2.6. Then the equalities*

$$\mathbf{\Pi}_r^\infty\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty\,, \tag{B.14}$$

$$\mathbf{\Pi}_r^\infty\,\mathbf{A}^{-1}\,\mathbf{E} = \mathbf{A}^{-1}\,\mathbf{E}\,\mathbf{\Pi}_r^\infty\,, \tag{B.15}$$

$$\left(\mathbf{A}^{-1}\,\mathbf{E}\right)^\nu\,\mathbf{\Pi}_r^\infty = \mathbf{0} \quad \textit{and} \quad \mathbf{\Pi}_l^\infty\,\left(\mathbf{E}\,\mathbf{A}^{-1}\right)^\nu = \mathbf{0} \tag{B.16}$$

are satisfied.

Proof. In order to show (B.14), the transformation

$$\begin{aligned} \tilde{\mathbf{A}} = \mathbf{P}\,\mathbf{A}\,\mathbf{Q} \quad \Leftrightarrow \quad \tilde{\mathbf{A}}^{-1} = \mathbf{Q}^{-1}\,\mathbf{A}^{-1}\,\mathbf{P}^{-1} \quad &\Rightarrow \mathbf{A}^{-1}\,\mathbf{P}^{-1} = \mathbf{Q}\,\tilde{\mathbf{A}}^{-1} \\ &\Rightarrow \mathbf{Q}^{-1}\mathbf{A}^{-1} = \tilde{\mathbf{A}}^{-1}\,\mathbf{P} \end{aligned} \tag{B.17}$$

is considered. This allows to write

$$\begin{aligned} \mathbf{\Pi}_r^\infty\,\mathbf{A}^{-1} &= \mathbf{Q}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix} \underbrace{\mathbf{Q}^{-1}\,\mathbf{A}^{-1}}_{\tilde{\mathbf{A}}^{-1}\,\mathbf{P}} = \mathbf{Q}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix} \underbrace{\begin{bmatrix} \mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}}_{\tilde{\mathbf{A}}^{-1}} \mathbf{P} = \mathbf{Q}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}\mathbf{P}\,, \\ \mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty &= \underbrace{\mathbf{A}^{-1}\,\mathbf{P}^{-1}}_{\mathbf{Q}\,\tilde{\mathbf{A}}^{-1}} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}\mathbf{P} = \mathbf{Q}\underbrace{\begin{bmatrix} \mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}}_{\tilde{\mathbf{A}}^{-1}} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}\mathbf{P} = \mathbf{Q}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_\infty} \end{bmatrix}\mathbf{P}\,, \end{aligned} \tag{B.18}$$

which proves $\mathbf{\Pi}_r^\infty\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty$.

To show (B.15) the relations from (B.14) and (B.2) can be used:

$$\mathbf{\Pi}_r^\infty\mathbf{A}^{-1}\,\mathbf{E} \overset{(B.14)}{=} \mathbf{A}^{-1}\,\mathbf{\Pi}_l^\infty\,\mathbf{E} \overset{(B.2)}{=} \mathbf{A}^{-1}\,\mathbf{E}\,\mathbf{\Pi}_r^\infty\,. \tag{B.19}$$

Finally (B.16) is proved through

$$\begin{aligned}\left(\mathbf{A}^{-1}\,\mathbf{E}\right)^{\nu}\mathbf{\Pi}_r^{\infty} &= \left(\mathbf{Q}\,\tilde{\mathbf{A}}^{-1}\,\mathbf{P}\,\mathbf{E}\,\mathbf{Q}\,\mathbf{Q}^{-1}\right)^{\nu}\mathbf{\Pi}_r^{\infty} = \mathbf{Q}\left(\tilde{\mathbf{A}}^{-1}\,\tilde{\mathbf{E}}\right)^{\nu}\mathbf{Q}^{-1}\mathbf{\Pi}_r^{\infty}\\ &= \mathbf{Q}\left(\begin{bmatrix}\mathbf{J}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_\infty}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{n_f} & \mathbf{0}\\ \mathbf{0} & \mathbf{N}\end{bmatrix}\right)^{\nu}\mathbf{Q}^{-1}\,\mathbf{Q}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_\infty}\end{bmatrix}\mathbf{Q}^{-1}\\ &= \mathbf{Q}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{N}^{\nu}\end{bmatrix}\mathbf{Q}^{-1} \overset{(\mathbf{N}^{\nu}=\mathbf{0})}{=} \mathbf{0}\,.\end{aligned} \tag{B.20}$$

The equality $\mathbf{\Pi}_l^{\infty}\left(\mathbf{E}\,\mathbf{A}^{-1}\right)^{\nu}$ can be shown accordingly. ■

Appendix C

Discussion of the Column Rank of $[\mathbf{E}\,\mathbf{V},\,\mathbf{B}]$

In the following the matrix $[\mathbf{E}\,\mathbf{V},\,\mathbf{B}]$, abbreviated with $\mathbf{K} := [\mathbf{E}\,\mathbf{V},\,\mathbf{B}] \in \mathbb{C}^{n\times(q+m)}$, is analyzed in the context of DAEs (i. e. for $\det(\mathbf{E}) = 0$). Since full column rank of $\mathbf{K}$ is a requirement for the CURE-framework[1], it is important to check this property during reduction (as done in Algorithm 4.4).

Aside from a numerical verification, general (analytical) conditions would help to get a better understanding of this issue and perhaps lead to a reduction scheme which guarantees this property. Unfortunately it seems to be difficult to find universally valid relations. One reason for this is that $\mathbf{V}$ represents the basis of an *accumulated* rational Krylov subspace, i. e. the column vectors belong to several independently specified rational Krylov subspaces. Up to the author's knowledge, there are no general statements about the column rank of $\mathbf{K}$ which would help to enforce this property during reduction so far. For this reason several thoughts related to this topic which arose during the work on this thesis are collected in the following.

First of all it is obvious, that $\mathbf{E}\,\mathbf{V} \in \mathbb{C}^{n\times q}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$ must have full column rank themselves. Considering that, the columns of $\mathbf{V}$ have to be linearly independent, which is assumed anyway (see Section 3.2). Note that a singular matrix $\mathbf{E}$ in the DAE-case does not imply rank deficiency of $\mathbf{K}$. This can be easily shown by the example

$$\left(\underbrace{\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}}_{\mathbf{E}}\underbrace{\begin{bmatrix}1\\ 0\end{bmatrix}}_{\mathbf{V}},\ \underbrace{\begin{bmatrix}0\\ 1\end{bmatrix}}_{\mathbf{B}}\right) = \underbrace{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}}_{\mathbf{K}} \quad\Rightarrow\quad \operatorname{rank}(\mathbf{K}) = 2 = q + m\,. \tag{C.1}$$

To obtain further relations, the special structures of the Weierstraß canonical form and the primitive basis $\mathbf{V}^{\mathrm{P}}$ are exploited. For this purpose let $\mathbf{P}$ and $\mathbf{Q}$ denote the transformation matrices as defined in Lemma 2.5 and $\mathbf{T}$ be as in Corollary 3.15 such that $\mathbf{V} = \mathbf{V}^{\mathrm{P}}\,\mathbf{T}$ holds. Moreover note that a multiplication of

$$\mathbf{A}\,\mathbf{V} - \mathbf{E}\,\mathbf{V}\,\mathbf{S}_V = \mathbf{B}\,\mathbf{R} \tag{C.2}$$

from the left with $\mathbf{P}$ leads to $\tilde{\mathbf{V}} = \mathbf{Q}^{-1}\,\mathbf{V}$ which is the unique solution of

$$\tilde{\mathbf{A}}\,\tilde{\mathbf{V}} - \tilde{\mathbf{E}}\,\tilde{\mathbf{V}}\,\mathbf{S}_V = \tilde{\mathbf{B}}\,\mathbf{R}\,. \tag{C.3}$$

[1] In particular full column rank of $\mathbf{K}$ is necessary for the equivalence of a parametrization of the reduced transfer function with $\mathbf{F}$ and $\mathbf{W}$ (see Theorem 4.20).

Using the definition of the primitive basis $\mathbf{V}^{\mathrm{P}}$

$$\mathbf{V}^{\mathrm{P}} = \Big[\dots,\, \mathbf{v}^{\mathrm{P}}_{ijk},\, \dots\Big]\,, \qquad \text{with } \mathbf{v}^{\mathrm{P}}_{ijk} = \Big[(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\mathbf{E}\Big]^{k-1}(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\mathbf{B}\,\mathbf{r}_{ij} \tag{C.4}$$

leads with $\tilde{\mathbf{V}} = \mathbf{Q}^{-1}\mathbf{V}^{\mathrm{P}}\,\mathbf{T}$ to

$$\tilde{\mathbf{V}} = \mathbf{Q}^{-1}\underbrace{\Big[\dots,\, \mathbf{v}^{\mathrm{P}}_{ijk},\, \dots\Big]}_{\mathbf{V}^{\mathrm{P}}}\mathbf{T} = \underbrace{\Big[\dots,\, \tilde{\mathbf{v}}^{\mathrm{P}}_{ijk},\, \dots\Big]}_{\tilde{\mathbf{V}}^{\mathrm{P}}}\mathbf{T}\,, \tag{C.5}$$

with $\tilde{\mathbf{V}}^{\mathrm{P}} = \mathbf{Q}^{-1}\mathbf{V}^{\mathrm{P}}$ and

$$\begin{aligned}
\tilde{\mathbf{v}}^{\mathrm{P}}_{ijk} &= \mathbf{Q}^{-1}\Big[(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\mathbf{E}\Big]^{k-1}(\mathbf{A}-s_i\,\mathbf{E})^{-1}\,\mathbf{B}\,\mathbf{r}_{ij}\\
&= \Big[\big(\tilde{\mathbf{A}}-s_i\,\tilde{\mathbf{E}}\big)^{-1}\,\tilde{\mathbf{E}}\Big]^{k-1}\big(\tilde{\mathbf{A}}-s_i\,\tilde{\mathbf{E}}\big)^{-1}\,\tilde{\mathbf{B}}\,\mathbf{r}_{ij}\\
&= \begin{bmatrix}\big(\mathbf{J}-s_i\,\mathbf{I}_{n_f}\big)^{-k}\,\tilde{\mathbf{B}}_f\\ \Big[(\mathbf{I}_{n_\infty}-s_i\,\mathbf{N})^{-1}\,\mathbf{N}\Big]^{k-1}(\mathbf{I}_{n_\infty}-s_i\,\mathbf{N})^{-1}\,\tilde{\mathbf{B}}_\infty\end{bmatrix}\mathbf{r}_{ij}\,.
\end{aligned} \tag{C.6}$$

According to [32, p. 9] the rank of a matrix is invariant concerning regular transformations (matrix equivalence) such that

$$\operatorname{rank}(\mathbf{K}) = \operatorname{rank}\left(\mathbf{P}\,\mathbf{K}\begin{bmatrix}\mathbf{T}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}\right) = \operatorname{rank}(\hat{\mathbf{K}})\,. \tag{C.7}$$

holds. Thus instead of $\mathbf{K}$ one can analyze $\hat{\mathbf{K}}$:

$$\begin{aligned}
\hat{\mathbf{K}} &= \mathbf{P}\,\mathbf{K}\begin{bmatrix}\mathbf{T}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix} = \Big[\mathbf{P}\,\mathbf{E}\,\mathbf{Q}\,\mathbf{Q}^{-1}\,\mathbf{V}\,\mathbf{T}^{-1},\, \mathbf{P}\,\mathbf{B}\Big] = \Big[\tilde{\mathbf{E}}\,\tilde{\mathbf{V}}^{\mathrm{P}},\, \tilde{\mathbf{B}}\Big]\\
&= \Big[\hat{\mathbf{k}}_{111},\, \dots,\, \hat{\mathbf{k}}_{ijk},\, \dots,\, \hat{\mathbf{k}}_{sr_sq_{sr_s}},\, \tilde{\mathbf{b}}_1,\, \dots,\, \tilde{\mathbf{b}}_m\Big]
\end{aligned} \tag{C.8}$$

with $\tilde{\mathbf{B}} = [\tilde{\mathbf{b}}_1,\, \dots,\, \tilde{\mathbf{b}}_m]$ and

$$\hat{\mathbf{k}}_{ijk} = \tilde{\mathbf{E}}\,\tilde{\mathbf{v}}^{\mathrm{P}}_{ijk} = \begin{bmatrix}\big(\mathbf{J}-s_i\,\mathbf{I}_{n_f}\big)^{-k}\,\tilde{\mathbf{B}}_f\\ \Big[\mathbf{N}\,(\mathbf{I}_{n_\infty}-s_i\,\mathbf{N})^{-1}\Big]^{k}\,\tilde{\mathbf{B}}_\infty\end{bmatrix}\mathbf{r}_{ij}\,. \tag{C.9}$$

Making use of $\hat{\mathbf{K}}$ the problem changes to: the matrix [$\mathbf{E}\,\mathbf{V}$, $\mathbf{B}$] has full column rank, if and only if all columns of $\hat{\mathbf{K}}$, i. e. $\hat{\mathbf{k}}_{ijk}$ and $\tilde{\mathbf{b}}_w$ for all i, j, k and $w \in \{1,\, \dots,\, m\}$, are linearly independent.

From (C.9) one can see, that it is sufficient, if the upper parts are linearly independent (since the column rank is considered). This is essential especially if the partitioning of the FOM into a strictly proper and improper subsystem is done by projection of $\mathbf{B}$, such that $\tilde{\mathbf{B}}^{\mathrm{sp}}_\infty = \mathbf{0}$. In this case the whole lower part of $\hat{\mathbf{K}}$ becomes zero.

Another interpretation is, that if an ODE is extended by a system of algebraic constraints (this corresponds to the lower part of $\hat{\mathbf{K}}$) to obtain a DAE, then the column rank of [$\mathbf{E}\,\mathbf{V}$, $\mathbf{B}$] does not decrease. Instead it might even rise if $\tilde{\mathbf{B}}_\infty \neq \mathbf{0}$.

Appendix D

The Generalized Resolvent Equation

The following relations were used to verify the analytic expressions of the gradient and Hessian of $\mathcal{J}$ during SPARK given in [30, p. 80] for the DAE-case (i. e. $\det(\mathbf{E}) = 0$). In order to shorten the notation, the abbreviations

$$\begin{aligned} \mathbf{A}_{s_1} &:= (\mathbf{A} - s_1\,\mathbf{E})\ , \\ \mathbf{A}_{s_2} &:= (\mathbf{A} - s_2\,\mathbf{E}) \end{aligned} \tag{D.1}$$

are used.

At first the *generalized resolvent* is defined:

Definition D.1 ([37, p. 18])**.** Let $\lambda\,\mathbf{E} - \mathbf{A}$ be a regular matrix pencil. Then the *generalized resolvent* is defined as

$$(\lambda\,\mathbf{E} - \mathbf{A})^{-1}\ , \qquad \text{with } \lambda \in \mathbb{C} \setminus \lambda(\mathbf{E},\, \mathbf{A})\ . \tag{D.2}$$

Using this, the *generalized resolvent equation* is formulated:

Lemma D.2 (adapted from [37, p. 18])**.** *Let $\lambda\,\mathbf{E} - \mathbf{A}$ be a regular matrix pencil. Then for all $s_1,\ s_2 \notin \lambda(\mathbf{E},\, \mathbf{A})$ the* generalized resolvent equation

$$\mathbf{A}_{s_1}^{-1} - \mathbf{A}_{s_2}^{-1} = (s_1 - s_2)\,\mathbf{A}_{s_2}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1} \tag{D.3}$$

holds.

In the following corollary additional relations are presented, which may be useful in similar investigations:

Corollary D.3. *Let $\lambda\,\mathbf{E} - \mathbf{A}$ be a regular matrix pencil. Then for all $s_1,\ s_2 \notin \lambda(\mathbf{E},\, \mathbf{A})$ the relations*

$$s_1\,\mathbf{A}_{s_1}^{-1} - s_2\,\mathbf{A}_{s_2}^{-1} = (s_1 - s_2)\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A}\,\mathbf{A}_{s_1}^{-1}\ , \tag{D.4}$$

$$(s_1 - s_2)\,\mathbf{A}_{s_2}^{-1}\,\mathbf{E}\left(\mathbf{A}_{s_1}^{-1} + \mathbf{A}_{s_2}^{-1}\right)\mathbf{E}\,\mathbf{A}_{s_1}^{-1} = \mathbf{A}_{s_1}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1} - \mathbf{A}_{s_2}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_2}^{-1}\ , \tag{D.5}$$

$$\mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} - \mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A} = (s_1 - s_2)\,\mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} \tag{D.6}$$

hold.

Proof. The first equation can be proved by multiplication with $\mathbf{A}_{s_2}$ from the left and $\mathbf{A}_{s_1}$ from the right:

$$\begin{aligned} s_1\,\mathbf{A}_{s_1}^{-1} - s_2\,\mathbf{A}_{s_2}^{-1} &= (s_1 - s_2)\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A}\,\mathbf{A}_{s_1}^{-1} \qquad \Big|\ \mathbf{A}_{s_2}\cdot\ \dots\ \cdot\mathbf{A}_{s_1} \\ s_1\,\underbrace{(\mathbf{A} - s_2\,\mathbf{E})}_{\mathbf{A}_{s_2}} - s_2\,\underbrace{(\mathbf{A} - s_1\,\mathbf{E})}_{\mathbf{A}_{s_1}} &= (s_1 - s_2)\,\mathbf{A} \\ (s_1 - s_2)\mathbf{A} - s_1\,s_2\,\mathbf{E} + s_2\,s_1\,\mathbf{E} &= (s_1 - s_2)\,\mathbf{A}\ . \end{aligned} \tag{D.7}$$

In order to show (D.5), (D.3) is inserted into the left hand side:

$$\begin{aligned} (s_1 - s_2)\,\mathbf{A}_{s_2}^{-1}\,\mathbf{E}\left(\mathbf{A}_{s_1}^{-1} + \mathbf{A}_{s_2}^{-1}\right)\mathbf{E}\,\mathbf{A}_{s_1}^{-1} &= \left(\mathbf{A}_{s_1}^{-1} - \mathbf{A}_{s_2}^{-1}\right)\mathbf{E}\,\mathbf{A}_{s_1}^{-1} + \mathbf{A}_{s_2}^{-1}\,\mathbf{E}\left(\mathbf{A}_{s_1}^{-1} - \mathbf{A}_{s_2}^{-1}\right) \\ &= \mathbf{A}_{s_1}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1} - \cancel{\mathbf{A}_{s_2}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1}} + \cancel{\mathbf{A}_{s_2}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1}} - \mathbf{A}_{s_2}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_2}^{-1} \\ &= \mathbf{A}_{s_1}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1} - \mathbf{A}_{s_2}^{-1}\,\mathbf{E}\,\mathbf{A}_{s_2}^{-1}\ . \end{aligned} \tag{D.8}$$

Finally (D.6) is proved by inserting (D.4) into the right hand side:

$$\begin{aligned} (s_1 - s_2)\,\mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} &= \mathbf{E}\left(s_1\,\mathbf{A}_{s_1}^{-1} - s_2\,\mathbf{A}_{s_2}^{-1}\right)\mathbf{E} \\ &= s_1\,\mathbf{E}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} - s_2\,\mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{E} + \underbrace{\mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} - \mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E}}_{=0} + \underbrace{\mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A} - \mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A}}_{=0} \\ &= -\underbrace{(\mathbf{A} - s_1\,\mathbf{E})}_{\mathbf{A}_{s_1}}\mathbf{A}_{s_1}^{-1}\,\mathbf{E} + \mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} + \mathbf{E}\,\mathbf{A}_{s_2}^{-1}\underbrace{(\mathbf{A} - s_2\,\mathbf{E})}_{\mathbf{A}_{s_2}} - \mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A} \\ &= -\mathbf{E} + \mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} + \mathbf{E} - \mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A} = \mathbf{A}\,\mathbf{A}_{s_1}^{-1}\,\mathbf{E} - \mathbf{E}\,\mathbf{A}_{s_2}^{-1}\,\mathbf{A}\ . \end{aligned} \tag{D.9}$$

■

Appendix E

List of Theorems and other Statements

Appendix F

References

[1] M. I. Ahmad and P. Benner. “Interpolatory Model Reduction Techniques for Linear Second-Order Descriptor Systems”. In: *European Control Conference*. Strasbourg, France, June 2014, pp. 1075–1079. ISBN: 978-3-9524269-1-3. DOI: 10.1109/ECC.2014.6862210.

[2] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Rice University, Houston, Texas: SIAM, 2005. ISBN: 978-0-89871-529-3. DOI: 10.1137/1.9780898718713.

[3] A. C. Antoulas, C. A. Beattie, and S. Gugercin. “Interpolatory model reduction of large-scale dynamical systems”. In: *Efficient Modeling and Control of Large-Scale Systems*. Ed. by J. Mohammadpour and K. M. Grigoriadis. Springer, 2010, pp. 3–58. ISBN: 978-1-4419-5756-6. DOI: 10.1007/978-1-4419-5757-3.

[4] A. C. Antoulas and D. C. Sorensen. “Approximation of large-scale dynamical systems: An overview”. In: *International Journal of Applied Mathematics and Computer Science* 11 (2001), pp. 1093–1121.

[5] A. Astolfi. “A new look at model reduction by moment matching for linear systems”. In: *IEEE Conference on Decision and Control* 46 (Dec. 2007), pp. 4361–4366. DOI: 10.1109/CDC.2007.4434367.

[6] C. Beattie and S. Gugercin. “A Trust Region Method for Optimal $\mathcal{H}_2$ Model Reduction”. In: *IEEE Conference on Decision and Control* 48 (Dec. 2009), pp. 5370–5375. DOI: 10.1109/CDC.2009.5400605.

[7] C. Beattie and S. Gugercin. “Realization-independent $\mathcal{H}_2$-approximation”. In: *IEEE Conference on Decision and Control* 51 (Dec. 2012), pp. 4953–4958. DOI: 10.1109/CDC.2012.6426344.

[8] D. J. Bender. “Lyapunov-Like Equations and Reachability/Observabiliy Gramians for Descriptor Systems”. In: *IEEE Transactions on Automatic Control* 32 (Apr. 1987), pp. 343–348. DOI: 10.1109/TAC.1987.1104589.

[9] P. Benner and T. Stykel. *Model Order Reduction for Differential-Algebraic Equations: a Survey*. Preprint MPIMD/15-19. Max Planck Institute Magdeburg, Nov. 2015. Available from http://www.mpi-magdeburg.mpg.de/preprints/.

[10] A. Castagnotto, M. Cruz Varona, L. Jeschek, et al. *sss & sssMOR: Analysis & Reduction of Large-Scale Dynamic Systems with MATLAB*. In Preparation. Technische Universität München.

[11] A. Castagnotto, H. K. F. Panzer, K.-D. Reinsch, et al. *Stability-Preserving, Adaptive Model Order Reduction of DAEs by Krylov-Subspace Methods.* Preprint arXiv: 1508.07227 [math.NA]. Technische Universität München, Aug. 2015. Available from http://arxiv.org/abs/1508.07227.

[12] K. E. Chu. "The Solution of the Matrix Equations $AXB - CXD = E$ AND $(YA - DZ, YC - BZ) = (E, F)$". In: *Linear Algebra and its Applications* 93 (Aug. 1987), pp. 93–105. DOI: 10.1016/S0024-3795(87)90314-4.

[13] C. Deng and Y. Wei. "Representations for the Drazin inverse of 2×2 block-operator matrix with singular Schur complement". In: *Linear Algebra and its Applications* 435 (Dec. 2011), pp. 2766–2783. DOI: 10.1016/j.laa.2011.04.033.

[14] D. S. Djordjevic and P. S. Stanimirovic. "On the Generalized Drazin Inverse and Generalized Resolvent". In: *Czechoslovak Mathematical Journal* 51 (Sept. 2001), pp. 617–634. DOI: 10.1023/A:1013792207970.

[15] K. Gallivan, A. Vandendorpe, and P. Van Dooren. "Model Reduction of MIMO Systems via Tangential Interpolation". In: *SIAM Journal on Matrix Analysis and Applications* 26 (Nov. 2004), pp. 328–349. DOI: 10.1137/S0895479803423925.

[16] G. H. Golub and C. F. Van Loan. *Matrix Computations.* 3rd ed. Baltimore: The Johns Hopkins University Press, 1996. ISBN: 978-0-801-85414-8.

[17] S. Gugercin, A.C. Antoulas, and C. Beattie. "$\mathcal{H}_2$ Model Reduction for Large-Scale Linear Dynamical Systems". In: *SIAM Journal on Matrix Analysis and Applications* 30 (June 2008), pp. 609–638. DOI: 10.1137/060666123.

[18] S. Gugercin, T. Stykel, and S. Wyatt. "Model Reduction of Descriptor Systems by Interpolatory Projection Methods". In: *SIAM Journal on Scientific Computing* 35 (Sept. 2013), B1010–B1033. DOI: 10.1137/130906635.

[19] A. Ilchmann and T. Reis. *Surveys in Differential-Algebraic Equations I.* Berlin Heidelberg: Springer-Verlag, 2013. ISBN: 978-3-642-34927-0. DOI: 10.1007/978-3-642-34928-7.

[20] A. Ilchmann and T. Reis. *Surveys in Differential-Algebraic Equations II.* Switzerland: Springer International Publishing, 2015. ISBN: 978-3-319-11049-3. DOI: 10.1007/978-3-319-11050-9.

[21] A. Ilchmann and T. Reis. *Surveys in Differential-Algebraic Equations III.* Switzerland: Springer International Publishing, 2015. ISBN: 978-3-319-22427-5. DOI: 10.1007/978-3-319-22428-2.

[22] D. Kressner. *Numerical Methods for General and Structured Eigenvalue Problems.* Berlin: Springer, 2005. ISBN: 978-3-540-24546-9. DOI: 10.1007/3-540-28502-4.

[23] P. Kunkel and V. Mehrmann. *Differential-Algebraic-Equations: Analysis and Numerical Solution.* Zürich: European Mathematical Society, 2006. ISBN: 3-03719-017-5. DOI: 10.4171/017.

[24] C. S. Lu. "Solution of the matrix equation $AX + XB = C$". In: *Electronics Letters* 7 (Apr. 1971), pp. 185–186. DOI: 10.1049/el:19710123.

[25] Lunze, J. *Regelungstechnik 2.* Berlin: Springer, 2014. ISBN: 978-3-642-53943-5. DOI: 10.1007/978-3-642-53944-2.

[26] J. Lunze. "Eigenschaften von linearen DAE-Systemen". In: *at - Automatisierungstechnik* 64 (Feb. 2016), pp. 81–95. DOI: 10.1515/auto-2015-0091.

[27] A. J. Mayo and A. C. Antoulas. “A framework for the solution of the generalized realization problem”. In: *Linear Algebra and its Applications* 425 (Mar. 2007), pp. 634–662. DOI: 10.1016/j.laa.2007.03.008.

[28] V. Mehrmann and T. Stykel. “Balanced Truncation Model Reduction for Large-Scale Systems in Descriptor Form”. In: *Dimension Reduction of Large-Scale Systems*. Ed. by P. Benner, V. Mehrmann, and D. C. Sorensen. Springer Berlin Heidelberg, 2005, pp. 83–115. ISBN: 978-3-540-24545-2. DOI: 10.1007/3-540-27909-1_3.

[29] V. Mehrmann and T. Stykel. “Descriptor Systems: A General Mathematical Framework for Modelling, Simulation and Control”. In: *at - Automatisierungstechnik* 54 (July 2006), pp. 405–415. DOI: 10.1524/auto.2006.54.8.405.

[30] H. K. F. Panzer. “Model Order Reduction by Krylov Subspace Methods with Global Error Bounds and Automatic Choice of Parameters”. Dissertation. Technische Universität München, Department of Mechanical Engineering, 2014. ISBN: 978-3-8439-1852-7.

[31] H. K. F. Panzer, S. Jaensch, T. Wolf, et al. “A Greedy Rational Krylov Method for $\mathcal{H}_2$-Pseudooptimal Model Order Reduction with Preservation of Stability”. In: *American Control Conference*. Washington, DC, USA, June 2013, pp. 5512–5517. ISBN: 978-1-4799-0178-4. DOI: 10.1109/ACC.2013.6580700.

[32] S. Roman. *Advanced Linear Algebra*. 3rd ed. New York: Springer, 2008. ISBN: 978-0-387-72828-5. DOI: 10.1007/978-0-387-72831-5.

[33] Rommes, J. and Kürschner, P. and Martins, N. and Freitas, F. D. *Power system examples on MOR Wiki*. May 2013. URL: http://morwiki.mpi-magdeburg.mpg.de/morwiki/index.php/Power_system_examples.

[34] W. Rudin. *Real and Complex Analysis*. 3rd ed. New York: McGraw-Hill, 1987. ISBN: 978-0-070-54234-1.

[35] M. Schmidt. “Systematic Discretization of Input/Output Maps and other Contributions to the Control of Distributed Parameter Systems”. Dissertation. Technischen Universität Berlin, Fakultät II - Mathematik und Naturwissenschaften, 2007. DOI: 10.14279/depositonce-1600.

[36] G. W. Stewart. “On the Sensitivity of the Eigenvalue Problem $Ax = \lambda Bx$”. In: *SIAM Journal on Numerical Analysis* 9 (Dec. 1972), pp. 669–686. DOI: 10.1137/0709056.

[37] T. Stykel. “Analysis and Numerical Solution of Generalized Lyapunov Equations”. Dissertation. Technische Universität Berlin, Fakultät II - Mathematik und Naturwissenschaften, 2002. DOI: 10.14279/depositonce-578.

[38] T. Stykel. “Gramian-Based Model Reduction for Descriptor Systems”. In: *Mathematics of Control, Signals, and Systems* 16 (Mar. 2004), pp. 297–319. DOI: 10.1007/s00498-004-0141-4.

[39] T. Stykel. “Low-rank iterative methods for projected generalized Lyapunov equations”. In: *Electronic Transactions on Numerical Analysis* 30 (2008), pp. 187–202. ISSN: 1068-9613.

[40] T. Stykel. “On some norms for descriptor systems”. In: *IEEE Transactions on Automatic Control* 51 (May 2006), pp. 842–847. DOI: 10.1109/TAC.2006.875010.

[41] P. Van Dooren, K. A. Gallivan, and P.-A. Absil. “$\mathcal{H}_2$-optimal model reduction of MIMO systems”. In: *Applied Mathematics Letters* 21 (Dec. 2008), pp. 1267–1273. DOI: 10.1016/j.aml.2007.09.015.

[42] T. Wolf. “$\mathcal{H}_2$ Pseudo-Optimal Model Order Reduction”. Dissertation. Technische Universität München, Department of Mechanical Engineering, 2014. ISBN: 978-3-8439-1926-5.

[43] T. Wolf, H. K. F. Panzer, and B. Lohmann. “$\mathcal{H}_2$ Pseudo-Optimality in Model Order Reduction by Krylov Subspace Methods”. In: *European Control Conference*. Zürich, Switzerland, June 2013, pp. 3427–3432. ISBN: 978-3-033-03962-9.

[44] T. Wolf, H. K. F. Panzer, and B. Lohmann. “Sylvester Equations and the Factorization of the Error System in Krylov Subspace Methods”. In: *Vienna Conference on Mathematical Modelling (MATHMOD)*. 2012.

[45] K. Yosida. *Functional Analysis*. Berlin: Springer, 1995. ISBN: 978-3-540-58654-8. DOI: 10.1007/978-3-642-61859-8.

[46] E. Zeidler, I. N. Bronstein, K. A. Semendjaew, et al. *Springer-Taschenbuch der Mathematik*. 3rd ed. Wiesbaden: Springer Vieweg, 2013. ISBN: 978-3-8351-0123-4. DOI: 10.1007/978-3-8348-2359-5.