Multirate Time Integration Methods for Efficient Time Domain Simulation of Large-Scale Systems and Differential-Algebraic Equations

Dissertation

zur Erlangung
des akademischen Grades
eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

der
Fakultät für Mathematik und Naturwissenschaften
der
Bergischen Universität Wuppertal (BUW)
vorgelegt von

Christoph Johannes Hachtel

geboren am 05.01.1987 in Leonberg

betreut von: Prof. Dr. Michael Günther (BUW)
PD Dr. Andreas Bartel (BUW)

Wuppertal, 7. September 2020

Contents

Abbreviations

DAE	Differential-Algebraic Equation
IVP	Initial Value Problem
LTI	Linear Time-Invariant System
MQS	Magneto Quasi-Static
MOR	Model Order Reduction
mrIRK*x*-DAE	Multirate Implicit Runge-Kutta Scheme for DAEs of Order x
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation

List of Symbols

Below, we present an alphabetically ordered list of all symbols, that are used in more than one section. Variables, that do not appear in this list, are defined locally and their definition is valid only for one section.

The denotation of physical quantities are chosen according to international standards. Therefore, the usage of one and the same letter for a physical quantity and a mathematical variable is inevitable during this thesis. However, within one section the relation between letters and their meaning is always injective.

Latin Letters

$A^{\ddagger}$ 14
System matrix of an LTI-system or linear ODE/DAE-system

$A^{\ddagger}_{SS}$, A_{FF} 22
System matrix of the slow or fast subsystem

A_{SF}, A_{FS} 22
Coupling-matrix between the slow and fast subsystem

a_{ij} 8
Coefficient of a Runge-Kutta method

$B^{\dagger,\ddagger}$ 14
Input matrix of an LTI-system

b_j 8
Weight of a Runge-Kutta method

$C^{\dagger,\ddagger}$ 14
Output matrix of an LTI-system

c_i 8
Stage of a Runge-Kutta method

E 29
An error

e 16
Euler's number, mathematical constant

$(\cdot)_F$ 5
Indicates the belonging to the fast subsystem

$\breve{F}^{\dagger}$, $\breve{G}^{\dagger}$ 74
Splitting functions of a dynamic iteration scheme

$f^{\dagger}$ 5
Function of the right hand side of an ODE or the differential part of a DAE

$G^{\dagger}$ 50
Implicit function of the algebraic constraints

$g^{\dagger}$ 49
Function of the right hand side of the algebraic part of a DAE

H 7
Macro-step size

h 7
Micro-step size

I 5
A time interval $[t_0, t_{\text{end}}]$

I_n 14
Identity matrix of dimension n

K_{ij} 19
Coupling matrix between the inputs and outputs of a coupled LTI-system

L_x^f 55
Lipschitz constant of the function f w.r.t. x

M 24
Mass matrix of an implicit ODE or a linear DAE

$M(\cdot,\cdot)$ 60
An M-Matrix of a special structure

m 7
Multirate factor $m = H/h$

P 16
Infinite Gramian matrix of reachability

Q 16
Infinite Gramian matrix of observability

r 14
Dimension of a reduced order system

$(\cdot)^r$ 14
Indicates the belonging to a reduced order system

$(\cdot)_S$ 5
Indicates the belonging to the slow subsystem

t 5
Time variable

$u(t)^\dagger$ 14
Input-function of an LTI-system

$W^\dagger$ 8
Intermediate stage value of the variable w in a Runge-Kutta scheme

$w^{\dagger,\ddagger}$ 5
The differential variable of an ODE/DAE

$\dot{w}^{\dagger,\ddagger}$, $\ddot{w}^\dagger$ 5
First and second time derivative of w

$w(t)^{\dagger,\ddagger}$ 5
The analytical solution of an ODE/DAE

$w_n^{\dagger,\ddagger}$ 7
The numerical approximation of an ODE/DAE

$x^\dagger$ 43
A variable, several definitions, given in context

$\dot{x}^{\dagger,\ddagger}$ 43
The time derivative of x

$y^\dagger$ 14
Output-function of an LTI-system

$\tilde{y}^\dagger$ 14
Output-function of a reduced-order LTI-system

$Z^\dagger$ 50
Intermediate stage value of the variable z in a Runge-Kutta scheme

$z^\dagger$ 49
The algebraic variable of a DAE

$\dot{z}^\dagger$, $\ddot{z}^\dagger$ 50
First and second time derivative of z

$z(t)^\dagger$ 49
The algebraic, analytical solution of a DAE

$z_n^\dagger$ 50
The algebraic, numerical approximation of a DAE

Greek Letters

α 75
Contractivity constant

γ 20
Balanced truncation error bound

Δ 54
Multirate local truncation error

δ . . . . . **??**
Single-rate local truncation error

ε 6
A positive constant

Λ 15
Diagonalmatrix of all eigenvalues of a matrix

λ 15
Eigenvalue of a matrix

μ 20
Logarithmic matrix norm

Φ 62
Differential update function

$\check{\Phi}$ 74
Extrapolation operator

Ψ 62
Algebraic update function

$\check{\Psi}$ 74
Dynamic iteration operator

ρ 19
Spectral radius

σ 17
Singular value of a matrix

Calligraphic and Other Letters

$\mathcal{E}$ 17
Rectangular matrix $\mathcal{E}_r = (I_r, O^\top)^\top$

$\mathcal{G}^\dagger$ 19
Transfer function of an LTI-system

$\mathcal{T}$ 14
Regular transformation matrix

$\mathcal{U}_\varepsilon(x)$ 52
Open environment of x with radius ε

$\mathcal{V}$ 14
Right MOR projection matrix

$\mathcal{W}$ 14
Left MOR projection matrix

$\mathbb{C}$ 19
The set of complex numbers

$\mathbb{N}$ 7
The set of natural numbers

$\mathbb{R}$ 5
The set of real numbers

$\mathbb{R}^+$ 15
The set of all positive real numbers

Physical Quantities

$\hat{A}$ 40
Magnetic vector potential

C 24
Capacity

$\mathfrak{e}$ 25
Energy

G 40
Electric conductivity

i 24
Electric current

P, p 25
Power

R 24
Electric resistance

T 24
Temperature

u 24
Electric Voltage

Λ, λ 24
Thermal conductivity

ν 40
Magnetic reluctivity

† These variables can be equipped with a subscript $(\cdot)_S$ or $(\cdot)_F$ to indicate their belonging to a slow or fast subsystem

‡ These variables can be equipped with a superscript $(\cdot)^r$ to indicate their belonging to a reduced order system

Chapter 1

Introduction

Currently, the development of any modern technical device depends strongly on efficient and reliable simulation tools. Simulation software allows one to change any property of a device and test the new configuration without manufacturing an expensive prototype. In general, we can assume that the complexity of a technical device increases the number of physical effects that have to be considered for the simulation. We illustrate this phenomenon by providing the following example of an integrated circuit: The smaller the device, the closer the conductor paths and this causes electromagnetic effects between the conductor paths which have to be considered beside the currents and voltages. Moreover, the performed energy leads to heating of the complete device so that thermal effects can no longer be neglected. Voltages and currents, electromagnetic effects and thermal heating are based on different physical laws which are described by different types of mathematical equations. For each type of equation, there also exist a class of specialised algorithms that guarantee an efficient and reliable simulation of the underlying physical effect. To consider several physical effects in the simulation, the mathematical equations have to be coupled to each other and the algorithms must be adapted according to the properties of the resulting, coupled system. In this work, we focus on coupled systems with a particular dynamical behaviour: Some subsystems provide high dynamic changes while the other ones change much slower. Algorithms for an efficient and reliable simulation in time domain for this kind of coupled systems are called multirate time integration schemes.

Multirate methods are characterized by integrating each subsystem with an individual chosen step size according to its dynamical behaviour. Whereas a classical integration scheme uses one global step size for all subsystems. Multirate time integration for ordinary differential equations (ODEs) were introduced in [Ric60]. Since then, many different multirate methods have been proposed in literature. Distinctive features are the basic integration scheme and the applied coupling approach. There exist for example multirate linear multistep methods [GW84], multirate Runge-Kutta methods [KR99, GK01, HS09, GS16], multirate Rosenbrock-Wanner or W-methods [GR93, Bar01, SHV07] and multirate extrapolation methods [EL97, CS10]. The coupling approach describes the communication between the subsystems on the different time grids during the simulation process. Established coupling approaches are Decoupled-Slowest-First [GW84, GR93], Coupled-Slowest-First [EL97, SHV07, HS09] and Coupled-First-Step [Ric60, KR99, GK01, Bar01, CS10]. This thesis is about the adaption and extension of multirate Runge-Kutta methods regarding the two following aspects: On the one hand, linear model order reduction for high dimensional subsystems and on the other hand, the numerical treatment of algebraic constraints in the subsystems.

The mathematical model of certain physical effects results in a high dimensional system of equations, for example the semi-discretisation of a partial differential equation (PDE) which describes thermal or electromagnetic effects. However, in addition the network approach to describe voltages and currents in an integrated electrical circuit usually leads to a large scale system of equations. The numerical treatment of such high dimensional systems is quite challenging concerning run time and memory requirements. To achieve a reliable simulation result in finite time, a model order reduction can be applied to the large scale system to compute a low dimensional replacement system [Ant05]. The dynamical behaviour of the system will not be affected by the model

order reduction. We consider a coupled system of at least one high dimensional subsystem to which we apply a model order reduction. We investigate the influence of the model order reduction of one subsystem on the run time of a multirate method. We propose a strategy to reformulate the subsystems, such that the computational effort of the multirate time integration of the coupled system with order reduced subsystem can be decreased significantly.

For the case of a linear, time-invariant system, an error bound for certain model order reduction techniques is available, that means that the quality of the approximation by the low order replacement system can be measured. In frequency domain, such an error bound is also given for coupled systems with order reduced subsystems [RS07]. We study the question, how does the model order reduction affect the approximation properties of the multirate method and derive an error bound in time domain.

The dynamical behaviour of many technical systems cannot be described only by ordinary differential equations since additional algebraic constraints have to be fulfilled. Such systems are called differential-algebraic equations (DAEs). By a DAE, the dynamical properties of mechanical multibody systems [ESF98] or voltages and currents in an electrical circuit [GF99] can be described. But also a space discretisation of a Maxwell's PDE to describe electromagnetic effects can lead to a DAE in case of non-conductive materials [Sch11]. The numerical treatment of DAEs is much more challenging than time integration of ODEs. This holds especially for multirate methods, since the algebraic constraints must be fulfilled during the evaluation of the coupling terms. Multirate methods for DAEs have been introduced in [BGK02]. Moreover, there exist two extensive works on multirate schemes for DAEs: In [Ver08] multirate schemes for DAEs based on BDF-methods are presented using a specialized stability analysis, in [Str06] a class of integration schemes based on the Coupled-First-Step approach is provided. In this thesis, we elaborate an order 1 multirate Runge-Kutta method for DAEs for all three coupling approaches and an order 2 scheme based on the Coupled-Slowest-First approach. The presented integration schemes are based on Runge-Kutta methods for ODEs and we can easily adapt this approach to derive further multirate Runge-Kutta methods for DAEs. The following convergence analysis is carried out in a straight forward way without using any stability definition. As a final result we see that for the chosen coupling strategy the maximum convergence order is 2.

Outline of this thesis

Here a brief summary of the chapters contained herein.

Chapter 2 – Preliminaries

We introduce the theory of multirate time integration of ordinary differential equations based on one-step methods. We start with the definition of a multiscale ODE, which represents the prototype for all ODEs that can be integrated efficiently by a multirate method. We give a formal definition of a multirate Runge-Kutta scheme for ODEs and discuss three established coupling approaches.

Chapter 3 – Multirate Time Integration for Order Reduced Systems

We apply a model order reduction to the slow changing subsystem and integrate the resulting, coupled system with a multirate method. We sketch the main concepts of model order reduction

for linear, time-invariant systems. We employ a multirate method to the coupled system with order reduced, slow subsystem and to the original system and compare the computation time. As benchmark example we choose an electrical circuit with a temperature dependent resistor. We show that the computation time can be decreased significantly if the dimension of the coupling interface is small. It follows the derivation of a combined error bound in time domain, which estimates the error caused by the model order reduction and by the multirate time integration. Finally, we present simulation results of a field-circuit coupled system, where the large scale DAE model of the electrical field is projected on a low order ODE system.

Chapter 4 – Multirate One-Step Methods for Differential-Algebraic Equations

This chapter deals with the derivation and analysis of multirate Runge-Kutta method for DAEs of index-1. After a short introduction into the theory and numerical treatment of DAEs, we present a multirate method for DAEs based on the implicit Euler scheme. We consider all three coupling approaches and show that the integration method has convergence order 1. Numerical tests confirm the previous theory by applying the scheme to a modified Prothero-Robinson equation and to a field-circuit coupled system. We proceed in a similar way to deduce an order 2 multirate method for DAEs from the LobattoIIIC scheme. At the end of this chapter, we combine the ideas of multirate time integration and dynamic iteration schemes to derive a proposition about the convergence of general multirate one-step method using the Decoupled-Slowest-First approach.

Chapter 5 – Summary

The thesis closes with Chapter 5. Here, we summarize all previous results regarding the application of model order reduction for multirate time integration and the efficient simulation of DAEs by multirate time integration. Furthemore, we point out the perspectives for future research based on this thesis.

Related scientific works

Several results of this thesis have been already published in [HBG16b] [HBG16a], [HBG$^+$18] and [HBGS19].

Chapter 2
Preliminaries

We start with the introduction of multirate schemes for multiscale ordinary differential equations based on Runge-Kutta methods. The here considered problem class are multiscale ordinary differential equations which are defined in Section 2.1. For an efficient time integration, multirate methods for multiscale ordinary differential equations are presented in Section 2.2. In this section, we focus on multirate methods based on Runge-Kutta schemes and discuss different coupling strategies. This chapter is mostly based on [GS20].

2.1 Multiscale Ordinary Differential Equations

The mathematical model of the dynamical behaviour of a physical, technical or other *real world* system, often leads to an initial value problem (IVP) of ordinary differential equations (ODEs)

$$\dot{w}(t) = \frac{d}{dt}w(t) = f(t,w(t)), \qquad w(t_0) = w_0 \tag{2.1}$$

with $t \in I = [t_0, t_{\text{end}}]$, $w: I \to \mathbb{R}^n$ and $f: I \times \mathbb{R}^n \to \mathbb{R}^n$ sufficiently smooth. The Lipschitz continuity of f with respect to w for all $t \in I$ guarantees the existence of a unique solution of the initial value problem.

The dynamical behaviour of a complex technical or physical system is often given on different timescales: Some parts of the system provide high dynamical changes, others are changing much slower. This particular behaviour is also described by the resulting mathematical model and leads to

Definition 1 (Multiscale ODE-IVP)**.** *An ODE-IVP* (2.1) *is called multiscale-ODE-IVP, if it can be written in one of the following forms:*

1. Multiscale Split ODE-IVP

$$\dot{w}(t) = f_F(t,w(t)) + f_S(t,w(t)), \qquad w(t_0) = w_0 \tag{2.2}$$

with $t \in I = [t_0, t_{end}]$, $w: I \to \mathbb{R}^n$ and f_F, $f_S: I \times \mathbb{R}^n \to \mathbb{R}^n$ sufficiently smooth. The fast dynamics of w are described by f_F, the slow ones by f_S.

2. Multiscale Partitioned ODE-IVP

$$\begin{pmatrix} \dot{w}_F(t) \\ \dot{w}_S(t) \end{pmatrix} = \begin{pmatrix} f_F(t, w_F(t), w_S(t)) \\ f_S(t, w_F(t), w_S(t)) \end{pmatrix}, \qquad \begin{pmatrix} w_F(t_0) \\ w_S(t_0) \end{pmatrix} = \begin{pmatrix} w_{F0} \\ w_{S0} \end{pmatrix} \tag{2.3}$$

with $t \in I = [t_0, t_{end}]$, $w_F: I \to \mathbb{R}^{n_F}$ the vector of all fast changing components, $w_S: I \to \mathbb{R}^{n_S}$ the vector of all slow changing components and corresponding, sufficiently smooth right hand sides $f_F: I \times \mathbb{R}^{n_F} \to \mathbb{R}^{n_F}$, $f_S: I \times \mathbb{R}^{n_S} \to \mathbb{R}^{n_S}$.

Analogously, we can define multiscale ODE-IVPs with more than two different time scales. For simplicity of notation, we restrict ourself to the here presented case of one fast changing subsystem and one slow changing subsystem. Each of the formulations in Definition 1 can be transformed into the other one:

- Given a partitioned ODE-IVP (2.3), we set

$$w(t) = \begin{pmatrix} w_F(t) \\ w_S(t) \end{pmatrix}, \quad \dot{w}(t) = \begin{pmatrix} f_F(t, w(t)) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ f_S(t, w(t)) \end{pmatrix}, \quad w(t_0) = \begin{pmatrix} w_F(t_0) \\ w_S(t_0) \end{pmatrix} \tag{2.4}$$

  to derive a split ODE-IVP.

- Given a split ODE-IVP (2.2), we set

$$\dot{w}_F(t) = f_F(t, w_F(t) + w_S(t)),\ \dot{w}_S(t) = f_S(t, w_F(t) + w_S(t)),\ w(t) := w_F(t) + w_S(t) \tag{2.5}$$

  and end up with a partitioned ODE-IVP.

We point out that the reformulation of the initial values is only well defined for the transformation from the partitioned form to the split form. For the other way round, any choice of w_{F0} and w_{S0} with $w_0 = w_{F0} + w_{S0}$ leads to a proper solution of (2.5).

Since both formulation are equivalent, we only consider partitioned ODE-IVPs (2.3) in this work. The following assumption is valid for the whole thesis:

Assumption 1. *We consider a multiscale partitioned ODE-IVP* (2.3)*. For a fixed positive constant* $\varepsilon > 0$ *and any induced matrix norm the following holds*

$$\left\| \frac{\partial f_S(t, w_F, w_S)}{\partial w_F} \right\| < \varepsilon \tag{2.6}$$

for all $t \in I$*.*

The assumption formalises that the the slow components w_S only depend weakly on the the fast components w_F. From a theoretical point of view on multiscale systems, it is possible to consider systems with a strong coupling from the slow variables w_S to the fast ones w_F, i.e. a large value for

$$\left\| \frac{\partial f_F(t, w_F, w_S)}{\partial w_S} \right\|.$$

The multiscale behaviour of such a system is maintained, but it cannot be integrated efficiently with a multirate time integration scheme due to stability issues. Therefore, we will assume a weak coupling between both subsystems later in the numerical part of this thesis.

To derive a numerical approximation of the multiscale ODE-IVP (2.3) with high accuracy, a fine time grid is necessary due to the high dynamical changes of the fast variables w_F. The small step sizes are applied to the complete ODE-IVP. Therefore, numerical integration of any multiscale ODE-IVP is quite costly, especially for systems with many slow components and only few fast ones, i.e. $n_S \gg n_F$. To derive a numerical approximation with high accuracy within a short computation time, multirate time integration methods are widely used.

2.2 Multirate Time Integration Methods

We recall the multiscale partitioned ODE-IVP (2.3) with $n_S \gg n_F$ and write it as two coupled subsystems

$$\dot{w}_F(t) = f_F(t, w_F(t), w_S(t)), \qquad w_F(t_0) = w_{F0} \tag{2.7}$$
$$\dot{w}_S(t) = f_S(t, w_F(t), w_S(t)), \qquad w_S(t_0) = w_{S0}. \tag{2.8}$$

We refer to (2.7) as *fast changing* or *fast* subsystem, and analogously to (2.8) as *slow changing* or *slow* subsystem. The analytical solution of (2.7) at time point $t_n \in I$ is denoted by $w_F(t_n)$ and $w_S(t_n)$ denotes the analytical solution of (2.8), respectively. A multirate time integration method computes a numerical approximation of (2.7-2.8) by using different step sizes for the subsystems according to their dynamical behaviour: The fast changing subsystem is integrated with a small *micro-step* size h, the slow changing subsystem is integrated with a larger *macro-step* size H. Often, there is a fixed ratio between micro- and macro-step size $m = H/h$ with $m \in \mathbb{N}$, called *multirate factor*. Figure 2.1 illustrates the integration strategy of a multirate method for one macro-step $t_n \to t_n + H$.

The basic integration scheme for the macro-step and the micro-steps are usually given by one-step or multistep methods. Multirate methods are discussed for the first time in [Ric60] using Runge-Kutta methods as basic integration scheme. In [GW84], multirate linear multistep methods are derived. Multirate methods based on extrapolation schemes are proposed in [EL97, CS10]. Linear implicit ROW- and W-methods are presented in [GR93, Bar01, BG02, SHV07]. Multirate Runge-Kutta methods have been derived and investigated in [KR99, Kvæ00, GK01, HS09]. In [GS16] the theory of multirate Runge-Kutta methods is generalised such that many previous approaches can be now reduced to a special case of an MGARK-scheme. Furthermore, MGARK-schemes can be applied to split multiscale ODE-IVPs (2.2), whereas the previous schemes are usually derived for partitioned multiscale ODE-IVPs.

For all mentioned multirate approaches, the crucial part is the coupling between the subsystems. That is, how to achieve the values of the fast subsystem during the integration of the slow subsystem and vice versa. We discuss several coupling approaches in Section 2.2.2, beforehand in Section 2.2.1, we give a formal defintion of multirate Runge-Kutta methods.

2.2.1 Multirate Runge-Kutta Schemes

To derive a mulitrate method based on a Runge-Kutta scheme, we consider a multiscale partitioned ODE-IVP (2.7-2.8) on the time interval $I = [t_0, t_{\text{end}}]$. We split the interval into N macro-steps of size H

$$t_0 < t_0 + H < \ldots < t_0 + NH = t_{\text{end}} \tag{2.9}$$

and each macro-step is split into m micro-steps of size h

$$t_n < t_n + h < \ldots < t_n + mh = t_n + H \tag{2.10}$$

with $t_n = t_0 + nH$ for $n = 0, \ldots, N-1$. For simplicity of notation, we restrict ourselves to a constant, global macro-step size H, the theory can be easily adapted to a step size controlled macro-step size. We assume, that the approximations $w_{Fn} \approx w_F(t_n)$, $w_{Sn} \approx w_S(t_n)$ at t_n for a fixed $n = 0, \ldots, N-1$ are already computed. For the integration of the slow changing subsystem (2.8),

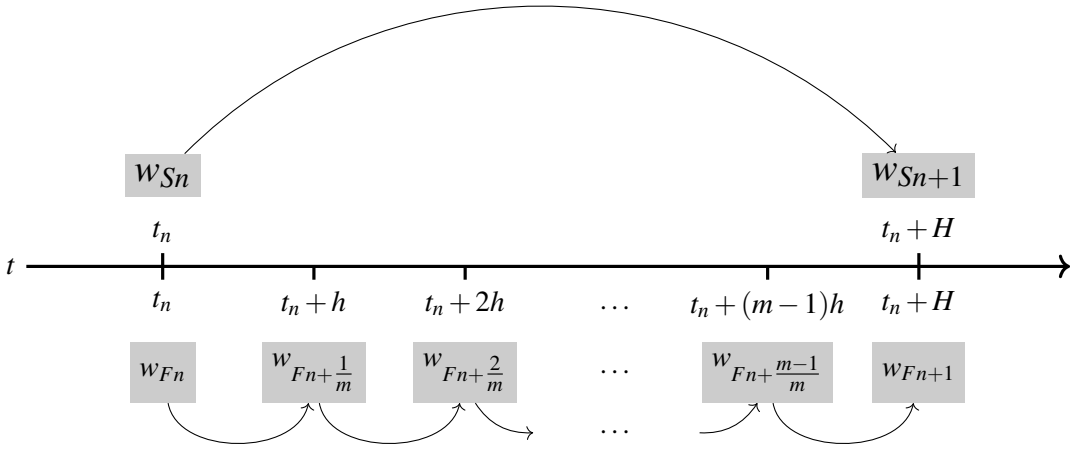


Figure 2.1: Schematic representation of a multirate time integration with constant micro-step size h: The slow changing variables w_S are integrated with one large macro-step size H, the fast changing variables w_F are integrated with m small micro-steps of size $h = H/m$. The numerical approximations are given by $w_{Sn} \approx w_S(t_n)$, $w_{Sn+1} \approx w_S(t_n + H)$ and possibly less accurate $w_{Fn+i/m} \approx w_F(t_n + ih)$, $i = 1, \dots m$, since the multirate scheme is interpreted as an integration method on the macro-step grid.

we consider an s-stage Runge-Kutta method with coefficient matrix $(a_{ij})_{i,j=1}^{s}$, stage vector $(c_i)_{i=1}^{s}$ and weight vector $(b_j)_{j=1}^{s}$. Then, the macro-step $t_n \to t_{n+1}$ for the slow changing subsystem reads

$$\begin{aligned} W_{Si} &= w_{Sn} + H \sum_{j=1}^{s} a_{ij} f_S(t_n + c_j H, \tilde{W}_{Fj}, W_{Sj}), \qquad i = 1, \dots, s, \\ w_{Sn+1} &= w_{Sn} + H \sum_{i=1}^{s} b_i f_S(t_n + c_i H, \tilde{W}_{Fi}, W_{Si}) \end{aligned} \tag{2.11}$$

with stage values $W_{Si} \approx w_S(t_n + c_i H)$. The values $\tilde{W}_{Fi} \approx w_F(t_n + c_i H)$ denote the coupling terms to the fast changing subsystems, there are several strategies to achieve these values, we will discuss them in Section 2.2.2. For the integration of the fast changing subsystem (2.8), we consider an $\tilde{s}$-stage Runge-Kutta method with coefficient matrix $(\tilde{a}_{ij})_{i,j=1}^{\tilde{s}}$, stage vector $(\tilde{c}_i)_{i=1}^{\tilde{s}}$ and weight vector $(\tilde{b}_j)_{j=1}^{\tilde{s}}$. During the macro-step $t_n \to t_{n+1}$, m micro-steps of size h are carried out. We assume that the approximation $w_{Fn+(l-1)/m} \approx w_F(t_n + (l-1)h)$ is already computed. Then, the micro-step $t_{n+(l-1)/m} \to t_{n+l/m} = t_n + lh$ reads

$$\begin{aligned} W_{Fi}^{l} &= w_{Fn+(l-1)/m} + h \sum_{j=1}^{\tilde{s}} \tilde{a}_{ij} f_F(t_n + \tilde{c}_j h, W_{Fj}^{l}, \tilde{W}_{Sj}^{l}), \qquad i = 1, \dots, \tilde{s}, \\ w_{Fn+l/m} &= w_{Fn+(l-1)/m} + h \sum_{i=1}^{\tilde{s}} \tilde{b}_i f_F(t_n + \tilde{c}_i h, W_{Fi}^{l}, \tilde{W}_{Si}^{l}) \end{aligned} \tag{2.12}$$

with stage values $W_{Fi}^{l} \approx w_F(t_n + (l-1)h + \tilde{c}_i h)$. The coupling terms to the slow changing subsystem are denoted by $\tilde{W}_{Si}^{l} \approx w_S(t_n + (l-1)h + \tilde{c}_i h)$.

The defintion allows to choose different Runge-Kutta methods for the macro-step and the micro-steps according to the properties of the subsystems (2.7) and (2.8). This multimethod approach is called *mixed-multirate* and was firstly derived in [Bar01] for linear implicit ROW-methods. In

case of a stiff, slow changing subsystem, an implicit Runge-Kutta method has to be used. Whereas the fast changing subsystem can be integrated by an explicit method by choosing the micro-step size sufficiently small. Here, one has to consider carefully the structure of the coupling between the subsystems: The coupled system (2.3) can provide other stiffness properties than the single subsystems (2.7) and (2.8).

In this work, we use one global Runge-Kutta method for the macro- and the micro-steps. The Runge-Kutta method is chosen according to the system properties of the coupled ODE-IVP (2.7-2.8). The stability of the resulting multirate methods also depends on the evaluation of the coupling variables $\tilde{W}_{Fi}$ in (2.11) and $\tilde{W}_{Si}^{l}$ in (2.12) which will be discussed in the following.

2.2.2 Coupling Strategies

Beside the choice of the underlying integration technique, the coupling strategy between the subsystems is the distinguishing element between different multirate methods.

In this work, we focus on *slowest-first* strategies: The integration process starts with the computation of an approximation of the slow subsystem to obtain an approximation of the slow subsystem $w_{Sn+1} \approx w_S(t_n + H)$, followed by the integration of the fast subsystem in the micro-steps. To evaluate the coupling term $\tilde{W}_{Fi}$ in the macro-step (2.11), there are three established strategies: Decoupled-Slowest-First, Coupled-Slowest-First and Coupled-First-Step. We introduce the different approaches, give the corresponding literature and compare the computational effort for the macro-step.

Subsequently, we discuss the evaluation of the coupling terms $\tilde{W}_{Si}^{l}$ in the micro-steps (2.12).

At the end of this section, we briefly introduce *fastest-first* strategies and explain the choice of using only *slowest-first* strategies in this work.

Decoupled-Slowest-First

The Decoupled-Slowest-First was proposed in [GW84] for multistep methods and in [GR93] for multirate ROW-methods. In the macro-step (2.11), the coupling variables $\tilde{W}_{Fi},\ i = 1,\dots,s$ are achieved by extrapolating the fast variables w_F. For multirate Runge-Kutta methods, one of the two following extrapolation formulas is usually used:

$$\text{Constant extrapolation:} \qquad W_{Fi} = w_{Fn}, \tag{2.13}$$

$$\text{Hermite extrapolation:} \qquad W_{Fi} = w_{Fn} + c_i h f_F(t_n, w_{Fn}, w_{Sn}) \tag{2.14}$$

for $i = 1,\dots,s$. The Hermite extrapolation formula corresponds to an integration step with the explicit Euler method for the fast subsystem (2.7). This often leads to stability issues in the time integration and makes the usage of constant extrapolation mandatory.

We point out, that the error of both extrapolation methods is of low order. This impacts also the resulting multirate method and limits its convergence order. To overcome this, a rational extrapolation method of higher accuracy is given in [GR93]. This method uses the Jacobian of f_F; either this information is given a-priori or its computation leads to additional effort.

In [Sch20] a promising alternative is presented: A cubic spline is computed, which interpolates

the approximations of w_F on the previous macro-step. This spline is evaluated at $t_n + c_i H$ to achieve the coupling term W_{Fi}.

Using a implicit Runge-Kutta method, the computational effort of the extrapolation is negligible compared to the computational effort of solving non-linear systems of equations. In case of a fully implicit Runge Kutte scheme, the Decoupled-Slowest-First approach leads to a non-linear systems of dimension $s \cdot n_S$, which has to be solved in each macro-step. To decrease a the dimension of the non-linear system, the fully implicit Runge-Kutta method can be replaced by a *Singly-Diagonal Implicit Runge-Kutta* (SDIRK) method [HNW08, Ch. II.7]. Here, s systems of linear equations of dimension n_S have to be solved in each macro-step.

Coupled-Slowest-First

For the macro-step in the Coupled-Slowest-First approach, both subsystems are solved simultaneously on the macro-step grid:

$$\begin{aligned} W_{Si} &= w_{Sn} + H \sum_{j=1}^{s} a_{ij} f_S(t_n + c_j H, W^*_{Fj}, W_{Sj}), \qquad i = 1, \ldots, s, \\ W^*_{Fi} &= w_{Fn} + H \sum_{j=1}^{s} a_{ij} f_F(t_n + c_j H, W^*_{Fj}, W_{Sj}), \qquad i = 1, \ldots, s, \\ w_{Sn+1} &= w_{Sn} + H \sum_{i=1}^{s} b_i f_S(t_n + c_i H, W^*_{Fi}, W_{Si}) \end{aligned} \tag{2.15}$$

with auxiliary stage values of the fast subsystem $W^*_{Fi} \approx w_F(t_n + c_i H),\ i = 1, \ldots, s$. Due to the large macro-step size, these approximations are inaccurate and therefore refused. This idea was introduced in [SHV07] for ROW-methods and in [HS09] for the θ-method.

The computation of the approximation w_{Sn+1} in (2.15) corresponds to the integration of the coupled ODE-IVP (2.3) with a classical single-rate Runge-Kutta scheme.

For the computation of w_{Sn+1} using the Coupled-Slowest-First approach, a system of non-linear equation of dimension $s \cdot (n_F + n_S)$ has to be solved in case of a fully implicit multirate Runge-Kutta method. Using a multirate method based on an SDIRK-scheme, s non-linear systems of dimension $n_F + n_S$ have to be solved in each macro-step.

Coupled-First-Step *or* Compound-Step-Methods

The Coupled-First-Step approach was introduced in [KR99] for Runge-Kutta schemes and in [Bar01] for ROW-methods as *Compound-Step* methods. The idea is, to compute the macro-step for the slow subsystem and the first micro-step for the fast subsystem simultaneously in one

compound step

$$
\begin{aligned}
W_{Si} &= w_{Sn} + H \sum_{j=1}^{s} a_{ij} f_S(t_n + c_j H, W_{Fj}^l, W_{Sj}), \qquad i = 1, \ldots, s, \\
W_{Fi}^1 &= w_{Fn} + h \sum_{j=1}^{s} a_{ij} f_F(t_n + c_j h, W_{Fj}^1, W_{Sj}), \qquad i = 1, \ldots, \tilde{s}, \\
w_{Sn+1} &= w_{Sn} + H \sum_{i=1}^{s} b_i f_S(t_n + c_i H, W_{Fi}^1, W_{Si}), \\
w_{Fn+1/m} &= w_{Fn} + h \sum_{i=1}^{s} b_i f_F(t_n + c_i h, W_{Fi}^1, W_{Si}).
\end{aligned}
$$

with stage values $W_{Si} \approx w_S(t_n + c_i H)$ and $W_{Fi}^1 \approx w_F(t_n + c_i h)$.

For a fully implicit Runge-Kutta scheme, the Coupled-First-Step approaches leads to a system of non-linear equations of dimension $s \cdot (n_F + n_S)$ for the macro-step and the first micro-step. In case of an SDIRK-method, s non-linear systems of dimension $n_F + n_S$ have to be solved.

After the macro-step or the compound-step, respectively, the integration process proceeds with the integration of the fast changing subsystem on the micro-step grid.

Micro-Steps

For the integration of the fast changing subsystem via (2.12) on the micro-step grid (2.10), the approximations of the slow changing subsystem $w_{Sn} \approx w_S(t_n)$, $w_{Sn+1} \approx w_S(t_{n+1})$ are available. Therefore, the coupling terms $\tilde{W}_{Si}^l \approx w_S(t_n + (l-1)h + \tilde{c}_i h)$ in (2.12) are achieved by one of the following interpolation formulas

$$
\begin{aligned}
\text{Constant interpolation at } t_n: &\qquad \tilde{W}_{Si}^l = w_{Sn} && (2.16)\\
\text{Constant interpolation at } t_{n+1}: &\qquad \tilde{W}_{Si}^l = w_{Sn+1} && (2.17)\\
\text{Linear Interpolation}: &\qquad \tilde{W}_{Si}^l = w_{Sn} + \frac{(l-1)+c_i}{m}\,(w_{Sn+1} - w_{Sn}) && (2.18)\\
\text{Hermite Interpolation}: &\qquad \tilde{W}_{Si}^l = w_{Sn} + (l-1+c_i)\; h \cdot f_S(w_{Fn}, w_{Sn}) && (2.19)
\end{aligned}
$$

for $i = 1, \ldots, s$ and $l = 1, \ldots, m$. Similarly to the extrapolation methods for the Decoupled-Slowest-First approach (2.13-2.14), the interpolation formulas (2.16-2.19) are of low accuracy and limit the convergence order of the resulting multirate method. In this work, we consider multirate methods of convergence order 1 and 2. We show that for these schemes, the accuracy of the above interpolation formulas is sufficient. To compute the coupling variables with a higher accuracy, a dense output formula can be applied [Bar01].

Considering the computational effort, we again neglect the interpolation of the slow changing variables for the computation of the coupling terms. To achieve an approximation $w_{Fn+1} \approx w_F(t_{n+1})$, m non-linear system of equations of dimension n_F have to be solved for the Decoupled-Slowest-First and the Coupled-Slowest-First approach. For the Coupled-First-Step approach, the computational effort for the remaining micro-steps reduces to $m-1$ non-linear systems of equations of dimension n_F.

Fastest-First-Strategies

For the Decoupled-Slowest-First approach, it is also possible to start the multirate scheme with the integration of the fast subsystem via (2.12) for $l = 1, \ldots, m$. Then, the values of the coupling terms $\tilde{W}_{Si}^{l} \approx w_S(t_n + (l-1)h + \tilde{c}_i h)$ are achieved by an extrapolation of the slow changing variables w_S. In the macro-step (2.11), the coupling terms $\tilde{W}_{Fi} \approx w_F(t_n + c_i H)$ are computed by an interpolation of the approximations of the fast subsystem $w_{Fn}, w_{Fn+1/m}, \ldots, w_{Fn+1}$. This procedure is called Decoupled-Fastest-First approach. We will not follow this approach since it might lead to a significant increase of computational effort when the multirate scheme is equipped with a step size control on the macro-step level [GW84]: If the error tolerance of the step size control is exceeded by the slow variables, also the micro-steps for the integration of the fast subsystem have to be recomputed based on a smaller macro-step size. Using a Slowest-First approach, the multirate factor m can be adapted if the error tolerance is exceeded by the fast subsystem without recomputing the macro-step for the slow variables.

Chapter Summary

This chapter provided the necessary mathematical knowledge on which the work of this thesis is based on. We introduced multiscale ordinary differential equations and discussed their efficient time integration with multirate methods. An overview of existing multirate methods in literature is given. We presented all established coupling approaches and derived in detail multirate Runge-Kutta methods.

In the next chapter, we will apply a projection based model order reduction to the slow changing subsystem of a multiscale ODE and investigate the impact on the computation time and the approximation properties of a multirate time integration method.

3 Chapter 3
Multirate Time Integration for Order Reduced Systems

The efficiency of a multirate time integration method compared to a classical single-rate integration scheme increases if the ratio $n_F : n_S$ between the number of fast and slow variables decreases. By using a multirate method, many function evaluations can be saved. In case of a very high dimensional slow subsystem, the main computational effort of the multirate time integration is still the evaluation of the slow changing subsystem. If the ODE-IVP is stiff and an implicit integration method has to be applied, a non-linear system of equations with many unknowns has to be solved, which decreases the performance of the integration method significantly. By applying a model order reduction (MOR) to the slow subsystem, the dimension n_S and therefore the number of slow changing variables can be reduced considerably, but the main properties of the slow subsystem are maintained [Ant05]. In [Ver08], the potential of multirate time integration and MOR is already discussed, but both concepts are considered separately. In this work, we are in particular interested in the interdependence of both concepts.

We start the chapter with a short introduction to linear MOR, to which we restrict ourselves in the following. Subsequent in Section 3.2, we will adapt the considered multiscale ODE-IVP such that a MOR can be applied to the slow subsystem and a the computation time of the multirate method can be decreased significantly. The impact of the MOR to the approximation properties of the integration method is investigated in Section 3.3. To this end, we derive an error bound in time domain that estimates the MOR caused error and the integration error of the multirate method. At the end of this chapter in Section 3.4, we present simulation results for a coupled field-circuit system where an MOR is applied to the electromagnetic field subsystem and the coupled system is integrated with an multirate method.

3.1 Model Order Reduction for Linear Time Invariant Systems

In this section, we present the basic idea of model order reduction (MOR) for high dimensional, dynamical systems. For further information we refer to [Ant05]. We focus on linear model order reduction for linear, time-invariant (LTI) systems and explain all theory which is necessary for the development of multirate time integration schemes with order reduced subsystems. We introduce projection based model order reduction for LTI-system (Section 3.1.1) and a particular model order reduction technique, namely balanced truncation (Section 3.1.2). In Section 3.1.3 we discuss briefly the specifics of model order reduction for coupled LTI-systems. Finally in Section 3.1.4, we repeat important properties of the logarithmic matrix norm.

3.1.1 Projection-Based Model Order Reduction

We start with an introduction to model order reduction for linear, time invariant systems (LTI). We consider an LTI-system in the following notation

$$\dot{w}(t) = Aw(t) + Bu(t), \quad w(t_0) = w_0 \tag{3.1}$$
$$y(t) = Cw(t) \tag{3.2}$$

with $t \in I = [t_0, t_{\text{end}}]$,a state space variable $w: I \to \mathbb{R}^n$, a system matrix $A \in \mathbb{R}^{n\times n}$, an input matrix $B \in \mathbb{R}^{n\times q}$, an external input function $u: I \to \mathbb{R}^q$, an output matrix $C \in \mathbb{R}^{p\times n}$ and an output function $y: I \to \mathbb{R}^p$. We assume that the dimension of the system n is very large. Such systems arise for example from a semi-discretisation of a partial differential equation (PDE). The aim of model order reduction is to find a reduced order LTI-system

$$\dot{w}^r(t) = A^r w^r(t) + B^r u(t), \quad w^r(t_0) = w_0^r \tag{3.3}$$
$$\tilde{y}(t) = C^r w^r(t) \tag{3.4}$$

with $w: I \to \mathbb{R}^r$, $A^r \in \mathbb{R}^{r\times r}$, $B^r \in \mathbb{R}^{r\times q}$, $C^r \in \mathbb{R}^{p\times r}$, $\tilde{y}: I \to \mathbb{R}^p$ and a reduced dimension $r \ll n$. The reduced order system is determined such that the input-output behaviour $u(t) \rightsquigarrow y(t)$ is approximated. This approximation property can be quantified by error in the output variable $\|y(t) - \tilde{y}(t)\|$ for $t \in I$. Usually, this error bound is given with respect to the L_2-norm. For some model order reduction techniques, this error can be bounded by a positive, real number ε.

$$\|y(t) - \tilde{y}(t)\| \leq \varepsilon. \tag{3.5}$$

Another reduction aim is the maintenance of certain properties of the original system by the reduced order system. One important property is the asymptotic behaviour of the state space variable:

Definition 2 (Asymptotic Stability)**.** *An LTI-sytem (3.1-3.2) is called asymptotic stable iff*

$$\lim_{t\to\infty} w(t) = 0 \quad \forall w_0. \tag{3.6}$$

or equivalently, if all eigenvalues of A have a negative real part.

To derive a reduced order system (3.3-3.4) the state space variable $w(t)$ is projected on an r-dimensional vector space. To this end, we consider a change of basis $\check{w} = \mathcal{T}w$ with a regular matrix $\mathcal{T} \in \mathbb{R}^{n\times n}$. We assume the following partitioning

$$\mathcal{T} = \begin{pmatrix} \mathcal{W}^\top \\ \mathcal{T}_1^\top \end{pmatrix} \qquad \text{and} \qquad \mathcal{T}^{-1} = (\mathcal{V},\ \mathcal{T}_2) \tag{3.7}$$

with $\mathcal{V}, \mathcal{W} \in \mathbb{R}^{n\times r}$. The transformed state space vector results in

$$\check{w} = \begin{pmatrix} \tilde{w} \\ \hat{w} \end{pmatrix}$$

for $\tilde{w} \in \mathbb{R}^r$ and $\hat{w} \in \mathbb{R}^{n-r}$. The model order reduction is realised by keeping $\tilde{w}$ and truncating $\hat{w}$. The defintion of $\mathcal{V}, \mathcal{W}$ in (3.7) leads directly to the relationship

$$\mathcal{W}^\top \mathcal{V} = I_r. \tag{3.8}$$

Therefore we can deduce that $\mathcal{V}\mathcal{W}^\top \in \mathbb{R}^{n\times n}$ is an oblique projection. We apply the truncated projection, performed by $\mathcal{V}$ and $\mathcal{W}$, to the LTI-system (3.1-3.2) and get

$$\dot{\tilde{w}}(t) = \mathcal{W}^\top A\mathcal{V}\tilde{w}(t) + \mathcal{W}^\top Bu(t), \quad \tilde{w}(t_0) = \mathcal{W}^\top w_0 \tag{3.9}$$

$$\tilde{y}(t) = C\mathcal{V}\tilde{w}(t) \tag{3.10}$$

which defines the reduced order LTI-system (3.3-3.4). There are several techniques how the projection matrices can be determined. A natural and simple approach for systems where the system matrix A is diagonalisable, one can compute the eigenvalue decomposition of A which reads

$$A = \mathfrak{V}\Lambda\mathfrak{V}^{-1}$$

with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ the matrix of all eigenvalues of A and $\mathfrak{V}$ the matrix of the corresponding eigenvectors. There are different approaches to sort the eigenvalues and eigenvectors. We follow the idea of the dominant pole approximation where the eigenvalues are arranged in descending order with respect to their real part $\Re(\lambda_1) > \Re(\lambda_2) > \ldots > \Re(\lambda_n)$. The reduced order model is now obtained by keeping the first r eigenvectors of A and truncate the remaining ones. To this end, we set in (3.7)

$$\mathcal{T} = \mathfrak{V}^{-1} = \begin{pmatrix} \mathfrak{w}_1 \\ \vdots \\ \mathfrak{w}_r \\ \mathcal{T}_1^\top \end{pmatrix}, \qquad \mathcal{T}^{-1} = \mathfrak{V} = (\mathfrak{v}_1, \ldots, \mathfrak{v}_r, \mathcal{T}_2)$$

and we have $\mathcal{V} = (\mathfrak{v}_1, \ldots, \mathfrak{v}_r)$ the first r columns of $\mathfrak{V}$ and $\mathcal{W} = (\mathfrak{w}_1^\top, \ldots, \mathfrak{w}_r^\top)$ the first r rows of $\mathfrak{V}^{-1}$. We point out that this method may fail if Jordan blocks of dimension greater than one arise in the eigenvalue decomposition of the system matrix A. Additionally, it stands out that the input and output matrices B and C of the LTI-system are not considered for the computation of the projection matrices $\mathcal{V}$ and $\mathcal{W}$. Usually, that leads to a worse approximation of the input-output behaviour of the reduced order system. Therefore, we now focus on a model order reduction technique which can be applied to a larger class of systems and which involves the matrices B and C. These requirements are fulfilled by balanced truncation model order reduction which is presented in the following section.

3.1.2 Balanced Truncation Model Order Reduction

The idea of balanced truncation model order reduction is that some states $w(t) \in \mathbb{R}^n$ of (3.1) are more important for the input-output behaviour than others. This is realised by keeping these important states and truncating the less important ones. To identify the relevant states, we consider the energy at the input and at the output of a certain state $\bar{w}$. To derive a proper energy measure, we have to introduce some concepts of system theory. For simplicity of notation, we consider in this section the time interval $I = [0, t_{\text{end}}]$. We start with a definition for the energy at the input of the LTI-system.

Definition 3 (Reachability)**.** *Let be given the LTI-system* (3.1) *with* $w(0) = 0$.

1. *A state* $\check{w} \in \mathbb{R}^n$ *is reachable, if there is a* $t^* \in \mathbb{R}^+$ *and a piecewise continuous input function* $u(t)$*, such that the solution* w *of* (3.1) *fulfils* $w(t^*) = \check{w}$.

2. *The Gramian matrix of reachability is defined for $t \in \mathbb{R}^+$ as*

$$P(t) = \int_0^t e^{A\tau} B B^\top e^{A^\top \tau} d\tau. \tag{3.11}$$

3. *The LTI-system* (3.1) *is called* fully reachable *if there is $t^* > 0$ such that $P(t^*)$ is positive definite.*

Now we can estimate the necessary energy of the input to reach a given state $\check{w}$ at time t^* by

$$\int_0^{t^*} u(t)^\top u(t) \mathrm{d}t \geq \check{w}^\top \left(P(t^*)\right)^{-1} \check{w}, \tag{3.12}$$

if $P(t^*)$ is regular. If the LTI-system is fully reachable, $P(t^*)$ is always regular. For the energy at the output of the LTI-system, we proceed similarly and derive the concept of

Definition 4 (Oberservability)**.** *Let be given the LTI-system (3.1-3.2).*

1. *A state $\hat{w}$ is unobservable, if for the particular input $u \equiv 0$, the solution $w(t)$ of* (3.1) *for the initial value $\hat{w} = w(0)$ always yields $y(t) = Cw(t) = 0, \quad \forall t \geq 0$.*

2. *The Gramian matrix of observability is defined for $t \in \mathbb{R}^+$ as*

$$Q(t) = \int_0^t e^{A^\top \tau} C^\top C e^{A\tau} d\tau. \tag{3.13}$$

3. *The LTI-system (3.1-3.2) is called* fully observable *if $Q(t)$ is positive definite for all $t \geq 0$.*

A measure for the energy at the output for the initial value $\hat{w} = w(0)$ and input $u \equiv 0$ is given by

$$\int_0^\infty y(t)^\top y(t) \mathrm{d}t = \hat{w}^\top Q \hat{w} \tag{3.14}$$

for $Q = \lim_{t\to\infty} Q(t)$, the infinite Gramian matrix of observability. Analogously, the infinite Gramian matrix of reachability is defined as $P = \lim_{t\to\infty} P(t)$. The definition of the energy of a certain state $\hat{w}$ at the input and the output of the system motivates the following reduction goals:

a) Truncate all states $\check{w}$ that are difficult to reach, i.e. states with a small value of

$$\check{w}^\top P \check{w} \qquad (\text{for} \quad \|\check{w}\| = 1).$$

b) Truncate all states $\hat{w}$ that are difficult to observe, i.e. states with a small value of

$$\hat{w}^\top Q \hat{w} \qquad (\text{for} \quad \|\hat{w}\| = 1).$$

For the particular case of a balanced LTI-system, both reduction goals are identical:

Definition 5 (Balanced System)**.** *An LTI-system (3.1-3.2) is called balanced, if the infinite Gramian matrices are given by*

$$P = Q = diag(\sigma_1, \ldots, \sigma_n), \quad with \quad \sigma_1 \geq \ldots \geq \sigma_n > 0, \tag{3.15}$$

with the Hankel singular values *of the the LTI-system* $\sigma_1, \ldots, \sigma_n$.

For a balanced system, the value

$$\bar{w}^\top P \bar{w} = \bar{w}^\top Q \bar{w} = \sigma_i$$

for $\bar{w} = e_i = (0, \ldots, 1, \ldots, 0)^\top$ indicates the importnace of the state $\bar{w}$.

For a general, asymptotic stable LTI-system with infinite Gramians $P,\ Q$, it is possible to apply a basis change $\mathcal{T}$, such that the transformed system is balanced. The infinite Gramians of the transformed system $\bar{P},\ \bar{Q}$ are given by

$$\bar{P} = \mathcal{T} P \mathcal{T}^\top, \qquad \bar{Q} = (\mathcal{T}^{-1})^\top Q \mathcal{T}^{-1}.$$

For balanced truncation model order reduction, these transformation matrices are computed, and the less important states of the balanced system are truncated. The detailed procedure is given in the following

Definition 6 (Balanced Truncation)**.** *Let be given a fully reachable, fully observable and asymptotic stable LTI-system (3.1-3.2) with infinite Gramian matrices P, Q.*

- $U \in \mathbb{R}^{n \times n}$ *denotes the Cholesky factor of P, i.e.* $P = UU^\top$.
- $K\Lambda^2 K^\top = U^\top Q U$ *denotes the eigenvalue decomposition with an orthogonal matrix* $K \in \mathbb{R}^{n \times n}$, $\Lambda = diag(\sigma_1, \ldots, \sigma_n)$ *and* $\sigma_1 \geq \ldots \geq \sigma_n > 0$.
- $\mathcal{T} := \Lambda^{1/2} K^\top U^{-1}$ *defines a regular transformation, such that the transformed LTI-system* $\bar{w} = \mathcal{T} w$ *is balanced and we have* $\bar{P} = \bar{Q} = \Lambda$.

Let $r \in \{1, \ldots, n-1\}$ *with* $\sigma_r > \sigma_{r+1}$ *and* $\mathcal{E}_r = \left(I_r,\ 0^\top\right)^\top$. *The balanced truncation projection matrices are given by*

$$\mathcal{V} := \mathcal{T}^{-1} \mathcal{E}_r, \qquad \mathcal{W}^\top := \mathcal{E}_r^\top \mathcal{T}.$$

The achieved reduced order system is again asymptotic stable, reachable, observable and it is balanced.

The existence of the Cholesky-decomposition $P = UU^\top$ follows from the positive definiteness of the matrix P and therefore from the reachability of the LTI-system. The entries $\sigma_1, \ldots, \sigma_n$ are exactly the Hankel-singular values of Definition 5. LTI-systems, which are not fully reachable can be projected on their *reachable subspace*. To this end, the singular value decomposition of P is computed

$$P = U_P \begin{pmatrix} \tilde{\Sigma}_P & 0 \\ 0 & 0 \end{pmatrix} V_P^*$$

with unitary matrices $U_P,\ V_P$ and a matrix $\tilde{\Sigma}_P = \mathrm{diag}(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_s)$ and $s = \mathrm{rk}(P)$. Then, the projec-

tion onto the reachable subspace is defined by

$$\mathcal{T}_R = U_P \mathcal{E}_S$$

with $\mathcal{E}_S = (I_S,\ 0^\top)^\top$. To the resulting, reachable subsystem the balanced truncation model order reduction can be applied.

A balanced truncation model order reduction requires the computation of the infinite Gramian matrices of the LTI-system. Usually, these matrices are not known a priori and have to be determined in advance. The computation of the limits of the matrix integrals in (3.11) and (3.13) is quite expensive and in most cases not possible. For LTI-systems of moderate dimension a common way is to solve the following equivalent Lyapunov-equations.

Lemma 2. *The infinite Gramian matrices P, Q of an asymptotic stable LTI-systems (3.1-3.2) are the unique solutions of the Lyapunov-equations*

$$AP + PA^\top + BB^\top = 0$$
$$A^\top Q + QA + C^\top C = 0.$$

For LTI-systems of high dimensions, alternative approaches to determine P and Q can be found in literature, for example based on ADI-schemes or Krylov-subspace methods [BS13, Sim16].

For balanced truncation, an a priori error bound is can be stated in terms of the Hankel singular values:

Theorem 3. *Let be given an asymptotic stable, fully observable and fully reachable LTI-sytem (3.1-3.2). The reduced order model (3.3-3.4) is achieved by balanced truncation (Def. 6). Then, the error in the output can be bounded by*

$$\|y - \tilde{y}\|_{L_2} \leq 2 \sum_{i=r+1}^{n} \sigma_i \|u\|_{L_2} \tag{3.16}$$

with σ_i the Hankel singular values of the LTI-system and r the dimension of the reduced order system.

For the particular choice of $C = I_n$ in (3.2) *and an input of unit energy, i.e. $\|u\|_{L_2} = 1$, the projection error is bounded by*

$$\|w - \mathcal{V}\mathcal{W}^\top w\|_{L_2} \leq 2(\sigma_{r+1} + \ldots + \sigma_n). \tag{3.17}$$

with w the solution of (3.1).

With balanced truncation we introduced a model order reduction method that preserves stability and provides an a priori error bound. In the next section, we extend the theory of model order reduction on coupled LTI-systems.

3.1.3 Model Order Reduction for Coupled LTI-systems

The main goal of the entire chapter is to apply a model order reduction to the slow changing subsystem of a multiscale ODE-IVP (2.3). The replacement of the slow subsystem by an order reduced system effects also the fast subsystem and the overall behaviour of the coupled system. Therefore, the theory of model order reduction has to be extended to coupled LTI-systems. In [RS07], the authors present the framework of k coupled LTI-system. To each of them, they apply a balanced truncation model order reduction and they investigate error bounds and stability properties for the coupled system. To keep it simple, we adapt the notation and the results to only two coupled systems and apply a model order reduction to only one of them.

We define the coupled, multiscale LTI-system as

$$\dot{w}_F(t) = A_F w_F(t) + B_F u_F(t), \tag{3.18}$$
$$y_F(t) = C_F w_F(t) \tag{3.19}$$
$$u_F(t) = K_{FF} y_F(t) + K_{FS} y_S(t) + H_F u(t), \tag{3.20}$$

$$\dot{w}_S(t) = A_S w_S(t) + B_S u_S(t), \tag{3.21}$$
$$y_S(t) = C_S w_S(t), \tag{3.22}$$
$$u_S(t) = K_{SF} y_F(t) + K_{SS} y_S(t) + H_S u(t), \tag{3.23}$$

with $t \in I = [t_0, t_{\text{end}}]$, state space variables $w_i : I \to \mathbb{R}^{n_i}$, system matrices $A_i \in \mathbb{R}^{n_i \times n_i}$, input matrices $B_i \in \mathbb{R}^{n_i \times q_i}$, output matrices $C_i \in \mathbb{R}^{p_i \times n_i}$, output functions $y_i : I \to \mathbb{R}^{p_i}$, global input matrices $H_i \in \mathbb{R}^{q_i \times q}$, a global input function $u : I \to \mathbb{R}^q$, coupling matrices $K_{ij} \in \mathbb{R}^{q_j \times p_i}$ for $i, j \in \{F, S\}$ and corresponding initial values $w_{F0} = w_F(t_0),\ w_{S0} = w_S(t_0)$. We assume that each subsystem itself is asymptotically stable. To derive a stability condition for the coupled LTI-system we introduce the transfer functions

$$\mathcal{G}_F(s) = C_F(sI_{n_F} - A_F)^{-1} B_F \tag{3.24}$$
$$\mathcal{G}_S(s) = C_S(sI_{n_S} - A_S)^{-1} B_S \tag{3.25}$$

for $s \in \mathbb{C} \backslash \lambda(A_i),\ i \in \{F, S\}$ with $\lambda(A_i)$ the set of all eigenvalues of A_i. The transfer function describes the input-output behaviour of an LTI-system in frequency domain and is a powerful tool in the analysis of model order reduction techniques. A norm for the transfer function is given by

$$\|\mathcal{G}\|_{\mathbb{H}_\infty} = \sup_{\omega \in \mathbb{R}} \|\mathcal{G}(i\omega)\|_2 \tag{3.26}$$

with respect to the spectral norm of $\mathcal{G}(i\omega)$ and the imaginary unit i. The asymptotic stability of the coupled LTI-system is characterised in the following Lemma.

Lemma 4. [RS07, Cor. 2.4] *Let be given the coupled LTI-system (3.18-3.23). The coupled LTI-system is asymptotic stable if the spectral radius* $\rho(\Pi) < 1$ *for*

$$\Pi = \tilde{K} \cdot diag(\|\mathcal{G}_F\|_{\mathbb{H}_\infty}, \|\mathcal{G}_S\|_{\mathbb{H}_\infty}), \qquad with \qquad \tilde{K} = \begin{pmatrix} \|K_{FF}\|_2 & \|K_{FS}\|_2 \\ \|K_{SF}\|_2 & \|K_{SS}\|_2 \end{pmatrix}. \tag{3.27}$$

We apply a balanced truncation model order reduction to the slow subsystem (3.21-3.22) such that the input-output behaviour $u_S(t) \rightsquigarrow y_S(t)$ is approximated. The output of the order reduced

subsystem is denoted by $\tilde{y}(t)$. The reduced order subsystem is of dimension r, is again asymptotic stable and the error bound is given by

$$\|y-\tilde{y}\|_{L_2} \leq \gamma \|u_S(t)\|_{L_2}$$

with $\gamma = 2(\sigma_{r+1} + \ldots + \sigma_{n_S})$ the truncated Hankel singular values of the subsystem. A stability condition for the coupled system with a reduced order subsystem is provided in the following

Theorem 5. [RS07, Theorem 4.1] *We consider the coupled LTI-system (3.18-3.23) with asymptotic stable subsystems. A balanced truncation model order reduction is applied to the subsystem (3.21-3.22) and the error bound is denoted by γ. The coupled LTI-system with order reduced subsystem is asymptotic stable if*

$$14 \cdot \gamma \cdot \|\tilde{K}\|_2 \|X\|_2 < 1 \tag{3.28}$$

with X the solution of the Lyapunov-equation

$$\Pi X \Pi^\top - X = -I_2 \tag{3.29}$$

and Π, $\tilde{K}$ *defined in* (3.27).

In [RS07] an error bound in frequency domain for the coupled LTI-system with reduced order subsystems is provided. We pursue another strategy and derive an error bound in time domain in Section 3.3.3. For the derivation of this error bound, some matrix theory from linear algebra is necessary. We provide this theory in the following section.

3.1.4 The Logarithmic Matrix Norm

The logarithmic matrix-norm for $A \in \mathbb{R}^{n\times n}$

$$\mu_x(M) := \lim_{h\to 0^+} \frac{\|I+hA\|_x - 1}{h} \tag{3.30}$$

is a powerful tool to analyse differential equations [Dah59]. Hereinafter, we will only use the logarithmic matrix norm for $x = 2$ and we skip the index in the following. For this particular case, the logarithmic matrix norm can be expressed by

$$\mu_2(A) := \lambda_{\max}\left(\frac{A+A^\top}{2}\right) \tag{3.31}$$

with $\lambda_{\max}$ the largest eigenvalue of $\frac{1}{2}(A+A^\top)$. The logarithmic matrix norm provides the following properties (amongst others):

$$\|e^{At}\|_2 \leq e^{\mu(A)t} \tag{3.32}$$

$$\mu(\mathcal{V}^\top A \mathcal{V}) \leq \mu(A) \tag{3.33}$$

for $t \in [t_0,\ t_{\text{end}}]$, any matrix $\mathcal{V} \in \mathbb{R}^{m\times r}$ with $\mathcal{V}^\top\mathcal{V} = I_r$ and $r < m$, [CS12].

Section Summary

We provided insight into the framework of projection based model order reduction for linear systems. With balanced truncation we introduced a model order reduction technique which maintains asymptotic stability of the system and which provides an error bound for the reduced order system. We pointed out that for coupled LTI-system additional conditions have to be fulfilled to guarantee stability also for order reduced subsystems. We will use the logarithmic matrix norm and its properties to derive a time domain error bound for a coupled system with order reduced subsystem. Equipped with this knowledge we continue to the heart of this chapter: multirate time integration with an order reduced, slow subsystem.

3.2 Model Order Reduction for Multiscale Ordinary Differential Equations

In the following, we combine the concepts of model order reduction and multirate time integration. Starting point is a partitioned multiscale initial value problem of ordinary differential equations in the form

$$\dot{w}_F(t) = f_F(w_F, w_S), \qquad w_F(t_0) = w_{F0} \tag{3.34}$$
$$\dot{w}_S(t) = f_S(w_F, w_L), \qquad w_S(t_0) = w_{S0} \tag{3.35}$$

providing the typical dynamical behaviour, such that the system can be integrated efficiently with a multirate time integration method: The components $w_F(t) \in \mathbb{R}^{n_F}$ are changing much faster than $w_S(t) \in \mathbb{R}^{n_S}$ and for the dimension it holds $n_F \ll n_S$. For simplicity of notation, but without loss of generality, we assume the coupled system to be autonomous. We assume a fixed and a-priori known partitioning of the subsystems according to their dynamic behaviour but the theory can be easily extended to more than two subsystems.

The multirate methods of this section are based on linear implicit one-step schemes which are briefly discussed in Section 3.2.1. In Section 3.2.2, we adapt the partitioned multiscale ODE (3.34-3.35) such that a model order reduction can be applied to the slow subsystem. The importance of a small dimensional coupling variables is stated in Section 3.2.3 followed by numerical results of a thermal-electric coupled system where the model order reduction is applied to a semi-discretised heat-equation (Section 3.2.4).

3.2.1 Multirate Linear Implicit Methods

We consider multirate linear implicit integration method with a Coupled-First-Step approach [BG02]. A simple integration scheme in that class is the multirate implicit Euler scheme [CS10]. The compound step for the linear implicit Euler method

$$\begin{pmatrix} I_{n_F} - h\dfrac{\partial f_F}{\partial w_F} & -\dfrac{h}{m}\dfrac{\partial f_F}{\partial w_S} \\ -mH\dfrac{\partial f_S}{\partial w_F} & I_{n_S} - H\dfrac{\partial f_S}{\partial w_S} \end{pmatrix} \begin{pmatrix} w_{Fn+1/m} - w_{Fn} \\ w_{Sn+1} - w_{Sn} \end{pmatrix} = \begin{pmatrix} hf_F(w_{Fn}, w_{Sn}) \\ Hf_L(w_{Fn}, w_{Sn})) \end{pmatrix} \tag{3.36}$$

and for the remaining micro steps holds

$$\left(I_{n_F} - h\,\frac{\partial f_F}{\partial w_F}\right) k_{F,i} = h\, f_F(w_{Fn+i/m}, \bar{w}_S(t_0+ih)), \qquad i = 1,\dots,m-1 \tag{3.37}$$

with $k_{F,i} = w_{Fn+(i+1)/m} - w_{Fn+i/m}$ and $\bar{w}_S$ the interpolated values of the slow components. This method can be interpreted as one Newton-iteration of the classic implicit Euler scheme and covers all properties of multirate linear implicit methods. It can be used to integrate moderately stiff systems of ODEs. The computational effort of one compound step consists of computing the $(n_F+n_S)\times(n_F+n_S)$-Jacobian matrix of the coupled and solving one linear system of equations of the same dimension.

3.2.2 Multirate Time Integration for Order Reduced Subsystems

In Section 2.2, we stated that multirate integration schemes exploit the special structure of the ODE (3.34-3.35) by using inherent time steps for the different subystems. Therefore, the high dimensional, slow subsystem has to be integrated considerably less often. Nevertheless, it remains to solve the high dimensional system which causes a large computational effort. This setting motivates to replace the large, slow subsystem of lower dimension, approximative replacement system. The idea is to compute this replacement system with a model order reduction as described in Section 3.1. To apply a linear model order reduction we assume that the slow subsystem (3.35) is linear-affine or can be linearised without loss of accuracy so that it can be written as

$$\dot{w}_S(t) = A_{SS} w_S(t) + (A_{SF},\ B_S) \begin{pmatrix} w_F(t) \\ u(t) \end{pmatrix}$$

with a system matrix $A_{SS} \in \mathbb{R}^{n_S \times n_S}$, a coupling matrix $A_{SF} \in \mathbb{R}^{n_S \times n_F}$, an input matrix $B_S \in \mathbb{R}^{n_S \times q_S}$ and an external, time dependent input $u\colon \mathbb{R} \to \mathbb{R}^{q_S}$. For the output of the slow changing subsystem $y_S(t) = C_S w_S(t)$ we set $C_S = I_{n_S}$. We apply a model order reduction to the slow variable w_S. To this end, the subsystem is projected on a low dimensional subspace by biorthogonal projection matrices $\mathcal{V}, \mathcal{W} \in \mathbb{R}^{n_S \times r}$ with $r \ll n_S$ as described in Section 3.1. In the fast subsystem (3.34) we replace $w_S(t)$ by $y_S(t)$ and end up with the reduced order, coupled system

$$\dot{\tilde{w}}_F(t) = f_F(\tilde{w}_F, y_S, t), \qquad \tilde{w}_F(t_0) = w_{F0} \tag{3.38}$$

$$\dot{w}^r_S(t) = \mathcal{W}^\top A_{SS} \mathcal{V} w^r_S(t) + \mathcal{W}^\top (A_{SF},\ B_S) \begin{pmatrix} \tilde{w}_F(t) \\ u_S(t) \end{pmatrix}, \qquad w^r_S(t_0) = \mathcal{W}^\top w_{S0} \tag{3.39}$$

$$y_S(t) = \mathcal{V} w^r_S(t). \tag{3.40}$$

$\tilde{w}_F$ is influenced by the model order reduction of the slow subsystem, but not reduced itself. Now, we apply the multirate linear implicit Euler method (3.36-3.37) to the order reduced, coupled system (3.38-3.40). For the upper-right off-diagonal block in the coefficient matrix in (3.36) we have

$$\frac{h}{m}\frac{\partial f_F}{\partial w^r_S} = \frac{h}{m}\frac{\partial f_F}{\partial y_S}\mathcal{V}$$

and it turns out that the coupling interface slow-to-fast is not reduced in this setting. In fact, with a non-reduced interface we cannot expect large improvements of the computational efficiency solving the system of linear equations in the compound step (3.36) by using a reduced, slow subsystem. So we have to find a way to transfer the reduced dimension to the coupling interface

to gain efficiency in the compound macro-step.

3.2.3 Interface Reduction

This section is based on results published in [HBG16a].

Often the fast components w_F do not depend on the detailed information of every single slow component. So we may replace the coupling interface w_S in (3.34) by a low dimensional input $u_F = g_F(y_F, y_S, u, t)$ with y_F, y_S the output of the subsystems and a global input u. The same procedure is made for the slow part (3.35). Adopting the notation for coupled linear systems from Section 3.1.3, we get for the full order model

$$\dot{w}_F = f_F(w_F, u_F, t) \qquad \dot{w}_S = f_S(w_S, u_S, t) := A_{SS} w_S + \hat{B}_S u_S \tag{3.41}$$

$$u_F = g_F(y_F, y_S, u, t) \qquad u_S = K_{SF} \cdot y_F + K_{SS} \cdot y_S + H \cdot u \tag{3.42}$$

$$y_F = h_F(w_F, t) \qquad y_S = C_S \cdot w_S \tag{3.43}$$

with subsystem dependent input $u_X(t) \in \mathbb{R}^{q_X}$, global input $u(t) \in \mathbb{R}^q$, output $y_X \in \mathbb{R}^{p_X}$, coupling matrices $K_{SF} \in \mathbb{R}^{n_S \times n_F}, K_{SS} \in \mathbb{R}^{n_S \times n_S}$, input matrices $\hat{B} \in \mathbb{R}^{n_s \times q_S}$, $H \in \mathbb{R}^{q_s \times q}$ and output matrix $C \in \mathbb{R}^{p_S \times n_S}$. We assume, that the reformulation of the coupled ODE-IVP (3.34-3.35) does not change its solution $w_F(t)$, $w_S(t)$ and we can keep the notation of variables from above.

The coupling functions g_F, h_F and matrices K_{SF}, C_S are not given by the system itself. Thus for the multirate setting they must be defined by the user exploiting some underlying properties, e.g. physical laws. These modifications in the multirate setting will not change the diagonal blocks in the compound step coefficient matrix (3.36), but for the off-diagonal blocks the mixed derivatives change into

$$\frac{\partial f_F}{\partial w_S} = \frac{\partial f_F}{\partial u_F} \cdot \frac{\partial g_F}{\partial w_S} = \frac{\partial f_F}{\partial u_F} \cdot \frac{\partial g_F}{\partial y_S} \cdot \frac{\partial y_S}{\partial w_S} = \frac{\partial f_F}{\partial u_F} \cdot \frac{\partial g_F}{\partial y_S} \cdot C_S. \tag{3.44}$$

$$\frac{\partial f_S}{\partial w_F} = \frac{\partial f_S}{\partial u_S} \cdot \frac{\partial u_S}{\partial w_F} = \frac{\partial f_S}{\partial u_S} \cdot \frac{\partial u_S}{\partial y_F} \cdot \frac{\partial y_F}{\partial w_F} = \hat{B}_S \cdot K_{SF} \cdot \frac{\partial h_F}{\partial w_F}. \tag{3.45}$$

On the right hand sides of (3.44) and (3.45) we find matrix products of the dimensions:

$$(n_F \times q_F) \cdot (q_F \times p_S) \cdot (p_S \times n_S) \tag{3.46}$$

$$(n_S \times q_S) \cdot (q_S \times p_F) \cdot (p_F \times n_F). \tag{3.47}$$

In a multirate context, the dimension n_F is supposed to be small. If the interface functions g_F, h_F, K_{SF}, C_S are chosen such that the dimension of their codomains are small, then only one large dimension remains, namely the number of the slow components n_S. However, we can compute a reduced model of dimension r for the slow part and use matrices $\hat{B}^r$ and C_S^r in the mixed derivatives of (3.44-3.45). Using this framework we expect higher efficiency in a time domain simulation.

3.2.4 Simulation of an Electric-Thermal Problem with Reduced Interface

To apply the theoretical considerations of the above sections, we use as benchmark example a modified version of the electric-thermal test circuit of [BGS03]: We deal with an electric circuit

Table 3.1: Parameters of the electric circuit

decide	parameter	decide	capacity
amplification	$A = 300$	capacity 1	$C_1 = 1F$
load resistance	$R_L = 0.3\text{k}\Omega$	capacity 2	$C_2 = 100\mu F$

pulsed voltage source $v(t) = \begin{cases} 0.5\sin(\pi t/(2.5\cdot 10^{-5}s))\ [\text{mV}] & \text{if} \quad t < 2.5\cdot 10^{-5}s \\ 0\ [\text{V}] & \text{otherwise} \end{cases}$

in which the thermal behaviour of a resistor is included. This results in a coupled system of the network equations and the heat equation. While voltages change very fast, heating or cooling of devices is a much slower process. Before applying the time integration, a semi-discretisation of space is performed for the heat equation. High accuracy demands as well as fine structures may lead to a large scale system. Therefore a model order reduction is applied to the slow, thermal subsystem. We work out the major ideas of the modelling process, for details see [BGS03].

Circuit Modeling. The electric part is represented by the circuit diagram in Fig. 3.1. The ODE model reads

$$C_1\dot{u}_3 = (u_2 - u_3)/R(T) - i_{\text{di}}(u_3 - u_4, T_{\text{di}}) \tag{3.48}$$
$$C_2\dot{u}_4 = i_{\text{di}}(u_3 - u_4, T_{\text{di}}) - u_4/R_L \tag{3.49}$$

with the node voltages u_3, u_4 and $u_2 = Av(t)$, the resistors's temperature T and the diode's temperature T_{di}. The nodal equations describe a quite stiff system of differential equations. So the multirate linear implicit Euler-method (3.36-3.37) is not a natural choice. To be able to apply this method to this circuit equations we use some extreme parameters amongst others for the capacitances. Table 3.1 shows all relevant parameters.

Between node two and three in the circuit, we consider a copper wire of length l and model it as a 1-D thermal dependent resistor. Let $a(x) = a_0 \cdot 1/(1 + (2/l)^2(l-x)x)$ denote the cross section of the wire while x represents the spatial coordinate; so at half of the length of the wire the cross section is half of the cross section at the ends. So we expect higher temperatures in the middle of the resistor. We assume a local resistance of the following type:

$$\mathfrak{r}(T) = r_0(1 + \tau(T - T_{\text{meas}})) \tag{3.50}$$

with thermal coefficient τ and specific resistance r_0 at temperature T_{meas}. We get the total resistance $R(T)$ by integrating the local resistance over the length of the wire l with respect to the cross section

$$R(T) = \int_0^l \frac{\mathfrak{r}(s, T(t,s))}{a(s)} ds = \int_0^l \tilde{\mathfrak{r}}(s, T(t,s)) ds. \tag{3.51}$$

The diode is also temperature dependent and has a strong non-linear behaviour, for the characteristic curve and more details see [BGS03].

Thermal Modeling and Coupling. The starting point of the thermal model is the 1-D heat equation for diffusive heat transport, which we use for the copper wire (resistor):

$$M_W'\dot{T} = \frac{\partial}{\partial x}\left(\Lambda(x)\frac{\partial T}{\partial x}\right) + \text{sources} \tag{3.52}$$

Figure 3.1: Circuit diagramm

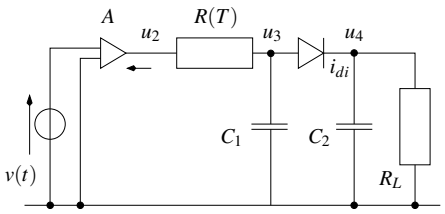
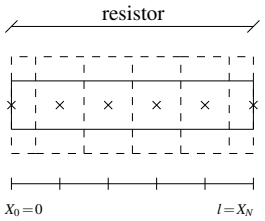


Figure 3.2: Finite Volume Discretised Resistor



with thermal mass of the wire M'_W and local 1-D conductivity $\Lambda(x) = \lambda(x) \cdot a(x)$. The sources term is comprised of two effects: (a) Local self heating due to the electric current. In fact, the dissipated power $P_W = u_R^2/R$ of the resistor results in heating the wire; (b) Cooling to the ambient temperature T_{env}, which is given by Newton's cooling $C = -\zeta S'(T - T_{\text{env}})$ with surface S'. For further details see [BGS03].

To be able to apply the multirate ODE-integration scheme (3.36-3.37), we discretise space in the parabolic PDE (3.52) by using the method of lines. We equip the wire with an equidistant grid $I_h : \ X_i = i \cdot k, \ i = 0, \ldots, N$ with $X_N = N \cdot k = l$ and use a finite volume approach. For that, we sub-divide the wire in cells of length k in the inner and $k/2$ at the boundaries. A schematic representation is given in Fig. 3.2. The heat conduction over one single cell can be simplified described by: *change of is inflow minus outflow*. So we get the approximation

$$M'_{W,i}\dot{T}_i = \Lambda \frac{T_{i+1} - 2T_i + T_{i-1}}{k^2} + P'_{W,i} - \zeta S'_{W,i}(T_i - T_{env}) \tag{3.53}$$

for the inner cells while i denotes the belonging of the variables to the i-th cell, $i = 1, \ldots, N-1$. For the boundary cells we have

$$M'_{W,0}\dot{T}_0 = \Lambda(T_1 - T_0)/k + P'_{W,0} - \zeta S'_{W,0}(T_0 - T_{env}) \tag{3.54}$$

$$M'_{W,N}\dot{T}_N = \Lambda(T_{N-1} - T_N)/k + P'_{W,N} - \zeta S'_{W,N}(T_N - T_{env}). \tag{3.55}$$

The diode is temperature dependent but without own thermal mass. So we just set the temperature at the end of the copper wire to be the temperature of the diode.

The coupling terms have been given indirectly in the models above:

(i) *Circuit to thermal:* Joule's law gives the dissipated power at the resistor. By adding an additional differential equation to the circuit equations,

$$\dot{\mathfrak{e}} = u_r \cdot i_r = (u_2 - u_3)^2/R(T), \tag{3.56}$$

the total energy $\mathfrak{e}$ is computed in each time step for the voltage u_r and the current i_r at the resistor $R(T)$. And $P_W = \mathfrak{e}/H^*$ gives us the required power for some time step H^*.

(ii) *Thermal to circuit:* Since the resistance $R(T)$ depends on the temperature profile T, we need the temperature distribution in the resistor to compute it, for a given distribution we use equation (3.51) to compute the total resistance. In addition, the diode's current depends on the wire temperature of the last cell.

We write the electric-thermal coupled system in the compact form

$$[\dot{u}_3, \dot{u}_4, \dot{\mathfrak{e}}] = f_F([u_3, u_4, \mathfrak{e}], T, t) \tag{3.57}$$

$$\dot{T} = f_S([u_3, u_4, \mathfrak{e}], T, t). \tag{3.58}$$

The computational cost of the compound step (3.36) depends on the number of discretisation points of the spatial variable of the thermal subsystem $\dim(T)$. If a high accuracy is demanded, this dimension can be large and the computational cost increases. So the question is how the coupling interface can be modified such that the dimension of the input of the fast part and the output of the slow part is small.

Reduced Interfaces. The heating of the resistor, caused by the electric current, is computed by the dissipated power p. The electric subsystem is computing the total dissipated energy $\mathfrak{e}$ in one macro step H. The ratio $\mathfrak{e}/H$ defines the averaged power, which we use for coupling [BGS03]. Hence we add an output function to the active subsystem: $h_F([u_3, u_4, \mathfrak{e}], t) = \mathfrak{e}/H$. To compute $\mathfrak{e}$, we have either to calculate differences of $\mathfrak{e}$ or we have to assign zero as the initial value for each macro step. If H is adjusted by a step size control, it has to be handled as an independent parameter.

For the coupling interface slow to active, one has to consider the thermal dependent, physical parameters, which are necessary in the circuit model and which can be computed by a linear model. In our case, these are the total resistance $R(T)$ and the diode's temperature T_{di}. Additional input functions for the slow and the active part are not necessary with this choice of coupling interfaces. As global input variable u we have the source voltage $v(t)$ which is used in the fast, electric subsystem only. These modifications in the interface of the coupled system (3.57-3.58) lead to

$$\begin{aligned} [\dot{u}_3, \dot{u}_4, \dot{\mathfrak{e}}] &= f_F([u_3, u_4, \mathfrak{e}], u_F, t) & \dot{T} &= A_{SS} \cdot T + B_S \cdot u_S \\ u_F &= [R(T), T_{di}, v(t)]^T & u_S &= p \\ p &= h_F([u_3, u_4, \mathfrak{e}], t) = \mathfrak{e}/H & [R(T), T_{di}] &= C_S \cdot T. \end{aligned}$$

For this system the off-diagonal blocks of the Jacobian matrix in the compound macro-step (3.36) become much smaller. Inspecting the dimensions like in (3.46) gives for $\frac{\partial f_F}{\partial w_S}$ the matrix sizes $(3 \times 2) \cdot (2 \times n_S)$ and for $\frac{\partial f_S}{\partial w_F}$ the dimension $(n_S \times 1) \cdot (1 \times 3)$. Now, a model order reduction can decrease the number of thermal variables from n_S to a significant smaller number r. No large dimensional terms occur in this setting so we expect a large gain concerning the computational effort using compound step multirate methods for this multiphysics application.

For the simulation of the system, we use the mixed multirate compound step method of [Bar01] which consists of a third order for the compound and a fourth order linear implicit method for the remaining micro steps. For the model order reduction we chose balanced truncation. We implemented the system and the integration methods in Matlab 2013a. All relevant simulation paramaters are listed in Table 3.2 and also the computation time can be seen there. The table shows the necessity of an interface reduction when combining a multirate scheme with a model order reduction: Only applying a model order reduction increases the computation time due to the loss of special matrix structures (sparsity, band structure). Interface reduction and MOR can decrease the computation time to 25%. Here, we are interested in two physical sizes: One is the temperature of the diode and the other is the highest temperature in the resistor which is found at its middle. Figure 3.4 shows the relative error of the multirate solution to the reference solution of these two physical sizes. Figure 3.3 shows the voltage curve at node three. The error is very

Figure 3.3: Voltage at Node 3

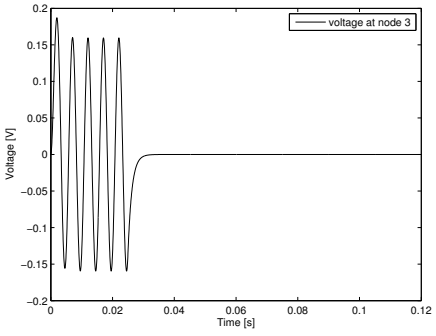


Figure 3.4: Relative errors

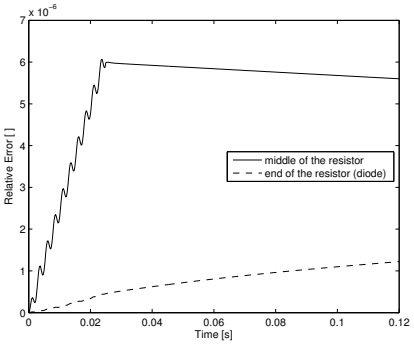


Table 3.2: Simulation parameters and computation time for full order model (FOM) and reduced order model (ROM) for simulation time $[0s,\ .12s]$

Model			**monolithic**		**interface reduced**	
Parameters	H	m	FOM $n_S = 50$	ROM $r = 5$	FOM $n_S = 50$	ROM $r = 5$
single-rate	$5 \cdot 10^{-5}$	1	$4.81s$	$5.65s$	$2.65s$	$1.91s$
multirate	$2.5 \cdot 10^{-4}$	5	$3.44s$	$5.00s$	$1.36s$	$1.18s$

small and we can say that our method decreases the computation time significantly with only a very small loss of accuracy.

Section Summary

We applied a model order reduction to the slow changing, linear subsystem of a partitioned multi-scale ODE-IVP and adapted the coupling interface between the subsystem, such that it is of small dimension. We applied a multirate time integration method to the resulting coupled system and observed a significant decrease of computational effort. The simulation results showed only a small loss of accuracy compared to the original, full order model. The investigation of the error that is caused by the model order reduction and its influence on the multirate time integration will be the topic of the next section.

3.3 An Error Estimate for the Multirate-MOR θ-Method

We study the impact of a model order reduction provided to one subsystem on the properties of a (multirate) time integration. To this end, we derive an error bound for a reduced order, coupled LTI-system in time domain and investigate consistency, stability and convergence of a multirate scheme applied to an order-reduced system.

3.3.1 Problem Setting

We consider a partitioned multiscale ODE-IVP (2.3) and assume both subsystems to be linear-affine. Then, the coupled system reads

$$\dot{w}_F(t) = A_{FF}w_F(t) + A_{FS}y_S(t) + u_F(t), \qquad w_F(t_0) = w_{F0} \tag{3.59}$$
$$\dot{w}_S(t) = A_{SS}w_S(t) + A_{SF}w_F(t) + B_S u_S(t), \qquad w_S(t_0) = w_{S0} \tag{3.60}$$
$$y_S(t) = C_S w_S(t) \tag{3.61}$$

with fast changing variables $w_F(t) \in \mathbb{R}^{n_F}$, slow changing variables $w_S(t) \in \mathbb{R}^{n_S}$, system matrices $A_{FF} \in \mathbb{R}^{n_F \times n_F}$, $A_{SS} \in \mathbb{R}^{n_S \times n_S}$, coupling matrices $A_{FS} \in \mathbb{R}^{n_F \times p_S}$, $A_{SF} \in \mathbb{R}^{n_S \times n_F}$, external input functions $u_F : [t_0, t_{\text{end}}] \to \mathbb{R}^{n_F}$, $u_S : [t_0, t_{\text{end}}] \to \mathbb{R}^{m_S}$, an input matrix $B_S \in \mathbb{R}^{n_S \times m_S}$, an output matrix $C_S \in \mathbb{R}^{p_S \times n_S}$ and the output of the slow subsystem $y_S(t) \in \mathbb{R}^{p_S}$. The coupling slow-to-fast is realised via an additional output function of the slow subsystem $y_S(t)$ and we can interpret the slow changing subsystem (3.60-3.61) as an LTI-system (3.1-3.2). The dimension of the slow changing subsystem is assumed to be much larger than the dimension of the fast subsystem $n_F \ll n_S$.

In context of model order reduction, we consider a slow subsystem with only few inputs and outputs, i.e. m_S, p_S are small, but with a large number of internal variables n_S. This property is formalised by writing the input and output of the system as a matrix-vector product $B_S u_S$ and $C_S w_S$, respectively. By contrast, we consider only a small number of internal variables for the fast subsystem n_F. Therefore, the input to the fast subsystem is written with a single input function u_F to keep the notation as simple as possible.

The coupled ODE system (3.59-3.61) can be rewritten in the following compact form

$$\underbrace{\begin{pmatrix} \dot{w}_F(t) \\ \dot{w}_S(t) \end{pmatrix}}_{=:\dot{w}(t)} = \underbrace{\begin{pmatrix} A_{FF} & A_{FS}C_S \\ A_{SF} & A_{SS} \end{pmatrix}}_{=:A} \begin{pmatrix} w_F(t) \\ w_S(t) \end{pmatrix} + \underbrace{\begin{pmatrix} u_F(t) \\ B_S u_S(t) \end{pmatrix}}_{u(t)}, \qquad w_0 = \begin{pmatrix} w_{F,0} \\ w_{S,0} \end{pmatrix}. \tag{3.62}$$

as one initial value problem for $w(t) := [w_F^\top(t),\ w_S^\top(t)]^\top$. Due to the multirate behaviour of the system, the coupling between the subsystems (3.59) and (3.60-3.61) is assumed to be weak. In this chapter, we formalise this property by the following

Assumption 6 (Weak Coupling)**.** *For the partitioned ODE system* (3.62) *we assume that the following holds*

$$\|A_{FS}C_S\| < \varepsilon \qquad \text{and} \qquad \|A_{SF}\| < \varepsilon \tag{3.63}$$

for an induced matrix norm and a fixed, small $\varepsilon \in \mathbb{R}^+$.

Furthermore, we assume that the system (3.62) is stable in sense of system theory:

Assumption 7 (Asymptotic Stability)**.** *The system matrix* $A = (a_{ij})_{i,j=1}^n$ *is strict diagonal dominant and its diagonal entries are all negative:*

$$a_{ii} + \sum_{i \neq j} |a_{ij}| < 0 \qquad \text{for } i = 1, \ldots, n. \tag{3.64}$$

We apply a projection based model order reduction as described in Section 3.1.1 to the slow changing subsystem (3.60-3.61), such that the input-output behaviour $[w_F(t),\ u_S(t)] \rightsquigarrow y_S(t)$ is

approximated and the number of internal state space variables $\dim(w_S^r)$ is small. The projection matrices are denoted by $\mathcal{V},\ \mathcal{W}\ \in \mathbb{R}^{n_S \times r}$. The coupled IVP with order reduced, slow subsystem reads

$$\dot{\tilde{w}}_F(t) = A_{FF}\tilde{w}_F(t) + A_{FS}\tilde{y}_S(t) + u_F(t), \qquad w_F(t_0) = w_{F0} \tag{3.65}$$

$$\dot{w}_S^r(t) = A_{SS}^r w_S^r(t) + A_{SF}^r \tilde{w}_F(t) + B_S^r u_S(t), \qquad w_S^r(t_0) = \mathcal{W}^\top w_{S0} \tag{3.66}$$

$$\tilde{y}_S(t) = C_S^r w_S^r(t) \tag{3.67}$$

with $w_S^r(t) \in \mathbb{R}^r$, $A_{SS}^r = \mathcal{W}^\top A_{SS}\mathcal{V}$, $A_{SF}^r = \mathcal{W}^\top A_{SF}$, $B_S^r = \mathcal{W}^\top B_S$ and $C_S^r = C_S\mathcal{V}$. The variables $\tilde{w}_F(t) \in \mathbb{R}_f^n$ and $\tilde{y}(t)_S \in \mathbb{R}^{p_S}$ are perturbed by the model order reduction but are not reduced themselves. To compute the projection matrices, we use balanced truncation model order reduction, cf. Section 3.1.2. The dimension r of the reduced order, slow subsystem is chosen such that the stability condition of Theorem 5 is fulfilled for

$$\tilde{K} = \begin{pmatrix} 0 & \|A_{FS}\|_2 \\ \|A_{SF}\|_2 & 0 \end{pmatrix}$$

and the balanced truncation error bound γ, cf. Theorem 3. Analogously to the compact form of the full order model (3.62) we rewrite the reduced order system (3.65-3.67) to

$$\begin{pmatrix} \dot{\tilde{w}}_F(t) \\ \dot{w}_S^r(t) \end{pmatrix} = \begin{pmatrix} A_{FF} & A_{FS}C_S^r \\ A_{SF}^r & A_{SS}^r \end{pmatrix} \begin{pmatrix} \tilde{w}_F(t) \\ w_S^r(t) \end{pmatrix} + \begin{pmatrix} u_F(t) \\ B_S^r u_S(t) \end{pmatrix}, \qquad w_0 = \begin{pmatrix} w_{F,0} \\ \mathcal{W}^\top w_{S,0} \end{pmatrix}. \tag{3.68}$$

To this system, we apply a multirate time integration scheme over the time interval $[t_0, t_{\text{end}}]$ with constant macro-step size H and fixed multirate factor $m \in \mathbb{N}$. The numerical approximation at t_n after n macro-steps is denoted by $[\tilde{w}_{Fn}^\top,\ w_{Sn}^{r\top}]^\top$ and y_{Sn}^r, respectively.

To derive a combined error bound for multirate time integration of coupled systems with order reduced, slow subsystems, we start with the definition of the estimated error.

3.3.2 Error Definition and Splitting

To investigate the impact of a model order reduction of the slow changing subsystem to the properties of a multirate time integration scheme, we are interested in the difference between the analytical solution of the full order system $w(t_n) \in \mathbb{R}^n$ and the numerical approximation of the reduced order system $[\tilde{w}_{Fn}^\top,\ w_{Sn}^{r\top}] \in \mathbb{R}^{n_F+r}$. Using the projection matrix $\mathcal{V}$ we can identify the numerical approximation of the reduced order system in the vector space $\mathbb{R}^n$. Then, the error reads

$$E(t) = \begin{pmatrix} E_F(t) \\ E_S(t) \end{pmatrix} = \begin{pmatrix} w_F(t) \\ w_S(t) \end{pmatrix} - \begin{pmatrix} \tilde{w}_{Fn} \\ \mathcal{V}w_{Sn}^r \end{pmatrix}. \tag{3.69}$$

In context of model order reduction, the projection matrices $\mathcal{V},\ \mathcal{W}$ are determined in such a way, that the input-output behaviour is approximated – or equivalently – the error in the output variable $\|y_S(t) - \tilde{y}_S(t)\|$ is minimized. For the particular choice of $C_S = I_{n_S}$ in (3.61), the error definition in (3.69) coincides with the output orientated error definition (3.5).

For the further investigation of the error (3.69), we split in the following way

$$\begin{aligned} E(t) &= \underbrace{\begin{pmatrix} w_F(t) \\ w_S(t) \end{pmatrix} - \begin{pmatrix} \tilde{w}_F(t) \\ \mathcal{V} w_S^r(t) \end{pmatrix}}_{E_{MOR}(t)} + \underbrace{\begin{pmatrix} \tilde{w}_F(t) \\ \mathcal{V} w_S^r(t) \end{pmatrix} - \begin{pmatrix} \tilde{w}_{Fn} \\ \mathcal{V} w_{Sn}^r \end{pmatrix}}_{E_{MR}(t)} \\ &= \qquad E_{MOR}(t) \qquad + \qquad E_{MR}(t) \end{aligned} \tag{3.70}$$

with $[\tilde{w}_F(t)^\top,\ w_S^r(t)^\top]^\top$ the analytical solution of the reduced order, coupled system (3.65-3.67). The error $E_{MOR}(t)$ is only caused by the model order reduction, while $E_{MR}(t)$ denotes the error of the time integration method applied to the *reduced order*, coupled system.

This splitting allows us to investigate the different errors with different techniques that fit for the error analysis of an MOR caused error and a time integration caused error, respectively.

According to the definition of E_{MR}, the multirate time integration scheme is only applied to the reduced order, coupled system (3.65-3.67). Therefore, the parameters of the integration method – macro-step size H and multirate factor m – are chosen such that the method converges and the simulation gives reliable results. An estimate for the error $E_{MR}(t)$ can be derived by using the results of the convergence analysis for multirate one-step methods for ODEs in terms of H, m and the Lipschitz constants of the reduced order, coupled system. Here, the multirate method is adapted to the reduced order, coupled system.

For other approaches, it can be interesting to turn around the setting, i.e. the parameters of the multirate method are adapted to the full order, coupled system and then, the model order reduction is applied in such a way, that the convergence properties of the integration method do not change. This problem setting is much more challenging and of less practical importance: If a reduced order model is already computed, the full order model is usually not integrated in time domain. For the theoretical point of view, we study this problem setting in Section 3.3.4: For the special case of a *balanced*, slow subsystem we can estimate the time integration error of the reduced order, coupled system in terms of the time integration error of the full order, coupled system using a perturbation argument of [HS09] for a multirate θ-method.

In Section 3.1.3 we discussed the existing results for error bounds and stability analysis for model order reduction of coupled systems. All these results have in common that the error is estimated in frequency domain. Using Parseval's theorem, e.g. [Gru19], the derived estimates can be transferred to time domain if L_2- and l_2-norms are used.

In combination with the time integration error E_{MR}, an estimate for E_{MOR} in frequency domain will not lead to a satisfying result for $E(t)$. Therefore, we provide an upper bound for the error E_{MOR} in time domain in the following section.

3.3.3 A Time-Domain Error Bound for E_{MOR}

Now we derive an estimate for the integral of the error $E_{MOR}(t)$ in (3.70). The error estimation is inspired by the work of Chaturantabut and Sorensen [CS12], who did a similar proof to derive a state-space error bound for POD-DEIM method. To investigate the error, we decompose it

according to the subsystems

$$\begin{aligned}\|E_{MOR}(t)\|_2^2 &= \left\| \begin{pmatrix} E_{F,MOR}(t) \\ E_{S,MOR}(t) \end{pmatrix} \right\|_2^2 = \|E_{F,MOR}(t)\|_2^2 + \|E_{S,MOR}(t)\|_2^2 \\ &\leq \|E_{F,MOR}(t)\|_2^2 + \|w_S(t) - \mathcal{V}w_S^r(t)\|_2^2 \\ &\leq \|E_{F,MOR}(t)\|_2^2 \\ &\quad + \| \underbrace{w_S(t) - \mathcal{V}\mathcal{W}^\top w_S(t)}_{=:\rho(t)} \|_2^2 + \| \underbrace{\mathcal{V}\mathcal{W}^\top w_S(t) - \mathcal{V}w_S^r(t)}_{=:\theta(t)} \|_2^2. \end{aligned} \tag{3.71}$$

Due to the coupling of the subsystems $E_{F,MOR}(t)$ and $\theta(t)$ depend on each other. Using Theorem 3 the error $\rho(t)$ can be estimated by

$$\int_0^{t_{\text{end}}} \|\rho(t)\|_2^2 dt \leq \gamma. \tag{3.72}$$

for a constant $\gamma > 0$. This error bound depends on the size of the reduced order, slow subsystem.

The crucial part for an estimation of the combined error $E_{MOR}(t)$ will be the estimation of $\theta(t)$. To this end, we define

$$\hat{\theta} = \mathcal{W}^\top \theta(t) = \mathcal{W}^\top w_S(t) - w_S^r(t)$$

due to the bi-orthogonality of the projection matrices $\mathcal{V}$ and $\mathcal{W}$ (3.8). Applying norms

$$\|\theta(t)\|_2 = \|\mathcal{V}\mathcal{W}^\top w_S(t) - \mathcal{V}w_S^r(t)\|_2 \leq \|\mathcal{V}\|_2 \|\hat{\theta}(t)\|_2,$$

we can estimate $\theta(t)$ by $\hat{\theta}(t)$. Using the differential equation of $w_S(t)$ and $w_S^r(t)$, we can derive for the differential equation for $\hat{\theta}(t)$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\hat{\theta}(t) &= \mathcal{W}^\top \left(A_{SS}w_S(t) + A_{SF}w_F(t) + B_S u_S(t)\right) - \left(A_{SS}^r w_S^r(t) + A_{SF}^r \tilde{w}_F(t) + B_S^r u_S(t)\right) \\ &= \mathcal{W}^\top A_{SS} w_S(t) - \mathcal{W}^\top A_{SS}\mathcal{V}w_S^r(t) + \mathcal{W}^\top A_{SF}\left(w_F(t) - \tilde{w}_F(t)\right) \\ &= \mathcal{W}^\top A_{SS} w_S(t) - \mathcal{W}^\top A_{SS}\mathcal{V}\mathcal{W}^\top w_S(t) + \mathcal{W}^\top A_{SS}\mathcal{V}\mathcal{W}^\top w_S(t) - \mathcal{W}^\top A_{SS}\mathcal{V}w_S^r(t) \\ &\quad + \mathcal{W}^\top A_{SF}E_{F,MOR}(t) \\ &= A_{SS}^r\left(\mathcal{W}^\top w_S(t) - w_S^r(t)\right) + \mathcal{W}^\top A_{SS}\left(w_S(t) - \mathcal{V}\mathcal{W}^\top w_S(t)\right) + \mathcal{W}^\top A_{SF}E_{F,MOR}(t) \\ &= A_{SS}^r\hat{\theta}(t) + \mathcal{W}^\top A_{SS}\rho(t) + \mathcal{W}^\top A_{SF}E_{F,MOR}(t).\end{aligned}$$

The analytical solution is given by

$$\hat{\theta}(t) = \int_0^t e^{A_{SS}^r(t-s)}\mathcal{W}^\top A_{SS}\rho(s)ds + \int_0^t e^{A_{SS}^r(t-s)}\mathcal{W}^\top A_{SF}E_{F,MOR}(s)ds, \tag{3.73}$$

since $\hat{\theta}(0) = \mathcal{W}^\top w_{S0} - \mathcal{W}^\top w_{S0} = 0$, see (3.68). In the same, way we write $E_{F,MOR}(s)$ as a differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}E_{F,MOR}(t) = A_{FF}E_{F,MOR}(t) + A_{FS}\left(\rho(t) + \theta(t)\right), \quad E_{F,MOR}(0) = 0 \tag{3.74}$$

and compute the analytical solution

$$E_{F,MOR}(t) = \underbrace{\int_0^t e^{A_{FF}(t-s)} A_{FS}\rho(s)ds}_{=:\tilde{\kappa}(t)} + \underbrace{\int_0^t e^{A_{FF}(t-s)} A_{FS}\theta(s)ds}_{=:\tilde{\varsigma}(t)}.$$

We define the functions $\kappa(t) := \mathcal{W}^\top A_{SF}\tilde{\kappa}(t)$ and $\varsigma(t) := \mathcal{W}^\top A_{SF}\tilde{\varsigma}(t)$. Applying norms to (3.73) and inserting the latest results we get the following inequality

$$\begin{aligned}\|\hat{\theta}(t)\|_2 \leq \|\mathcal{W}^\top A_{SS}\|_2 \int_0^t \|e^{A^r_{SS}(t-s)}\|_2 \|\rho(s)\|_2 ds \\ + \int_0^t \|e^{A^r_{SS}(t-s)}\|_2 \|\kappa(s)\|_2 ds + \int_0^t \|e^{A^r_{SS}(t-s)}\|_2 \|\varsigma(s)\|_2 ds.\end{aligned} \tag{3.75}$$

Our next step is to estimate the involved terms:

$$\|\kappa(t)\|_2 \leq \|\mathcal{W}^\top A_{SF}\|_2 \|A_{FS}\|_2 \int_0^t \|e^{A_{FF}(t-s)}\|_2 \|\rho(s)\|_2 ds.$$

To the integral, we apply Cauchy-Schwarz inequality and estimate the matrix exponential by the logarithmic matrix norm (3.32). Theorem 3 and setting $\mu_F := \mu(A_{FF})$ lead to

$$\|\kappa(t)\|_2 \leq \gamma \|\mathcal{W}^\top A_{SF}\|_2 \|A_{FS}\|_2 \left(\int_0^t (e^{\mu_F(t-s)})^2 ds \right)^{1/2}.$$

We introduce following short hand notation

$$q_\mu(t) = \int_0^t e^{\mu(t-s)} ds = \begin{cases} \frac{1}{\mu}(e^{\mu t} - 1) & \mu \neq 0 \\ t & \mu = 0 \end{cases} \tag{3.76}$$

and conclude the estimation of $\kappa(t)$ with

$$\|\kappa(t)\|_2 \leq \underbrace{\gamma \|\mathcal{W}^\top A_{SF}\|_2 \|A_{FS}\|_2 \left(q_{2\mu_F}(t_{\text{end}})\right)^{1/2}}_{:=\varphi}.$$

For the second summand in (3.75), we use the latest result and arguments from above to achieve

$$\int_0^t \|e^{A^r_{SS}(t-s)}\|_2 \|\kappa(s)\|_2 ds \leq \gamma\varphi \int_0^t \|e^{A^r_{SS}(t-s)}\|_2 ds \leq \gamma\varphi \int_0^{t_{\text{end}}} e^{\mu_S(t_{\text{end}}-s)} ds = \gamma\varphi q_{\mu_S}(t_{\text{end}})$$

with $\mu_S = \mu(A^r_{SS})$.

For the first summand in (3.75) we get analogously

$$\|\mathcal{W}^\top A_{SS}\|_2 \int_0^t \|e^{A^r_{SS}(t-s)}\|_2 \|\rho(s)\|_2 ds \leq \gamma \|\mathcal{W}^\top A_{SS}\|_2 \left(q_{2\mu_S}(t)\right)^{1/2}.$$

The function $\varsigma(t)$ can be estimated in terms of $\theta(t)$ as follows:

$$\begin{aligned}\|\varsigma(t)\|_2 &= \|\mathcal{W}^\top A_{SF}\|_2 \|A_{FS}\|_2 \int_0^t \|e^{A_{FF}(t-s)}\|_2 \|\theta(s)\|_2 ds \\ &\leq \|\mathcal{W}^\top A_{SF}\|_2 \|A_{FS}\|_2 \left(q_{2\mu_F}(t_{\text{end}})\right)^{1/2} \left(\int_0^t \|\theta(s)\|_2^2 ds\right)^{1/2} \\ &= \varphi \left(\int_0^t \|\theta(s)\|_2^2 ds\right)^{1/2}.\end{aligned}$$

With this result, we get for the last summand in (3.75):

$$\begin{aligned}\int_0^t \|e^{A_{SS}^r(t-s)}\|_2 \|\varsigma(s)\|_2 ds &\leq \varphi(q_{2\mu_S}(t_{\text{end}}))^{1/2} \left(\int_0^t \int_0^s \|\theta(\tau)\|_2^2 d\tau ds\right)^{1/2} \\ &\leq \varphi(q_{2\mu_S}(t_{\text{end}}))^{1/2} \left(\int_0^t \int_0^t \|\theta(\tau)\|_2^2 d\tau ds\right)^{1/2} \\ &\leq \varphi(q_{2\mu_S}(t_{\text{end}}))^{1/2} \left(t \int_0^t \|\theta(\tau)\|_2^2 d\tau\right)^{1/2}. \qquad (3.77)\end{aligned}$$

Inserting all estimations into (3.75), we find

$$\|\theta(t)\|_2 \leq \|\mathcal{V}\|_2 \|\hat{\theta}(t)\|_2 \leq \|\mathcal{V}\|_2 \left(\eta + \varphi(q_{2\mu_S}(t_{\text{end}}))^{1/2} \left(t \int_0^t \|\theta(\tau)\|_2^2 d\tau\right)^{1/2}\right) \qquad (3.78)$$

with

$$\eta := \gamma \|\mathcal{W}^\top A_{SS}\|_2 \left(q_{2\mu_S}(t)\right)^{1/2} + \gamma\varphi q_{\mu_S}(t_{\text{end}}).$$

Building the square of the previous inequality (3.78) and using that $(a+b)^2 \leq 2a^2 + 2b^2$, we obtain

$$\|\theta(t)\|_2^2 \leq \underbrace{2\|\mathcal{V}\|_2^2 \eta^2}_{=:\hat{\eta}} + \underbrace{2\|\mathcal{V}\|_2^2 \varphi^2 q_{2\mu_S}(t_{\text{end}}) t_{\text{end}}}_{=:\hat{\varphi}} \int_0^t \|\theta(\tau)\|_2^2 d\tau = \hat{\eta} + \hat{\varphi} \int_0^t \|\theta(\tau)\|_2^2 d\tau.$$

Applying Gronwall's lemma, we get

$$\|\theta(t)\|_2^2 \leq \hat{\eta} e^{\hat{\varphi} t}. \qquad (3.79)$$

With this result, we can estimate the error in the fast subsystem for $E_{F,MOR}$, starting with equation (3.74) as follows:

$$\|E_{F,MOR}(t)\|_2^2 \leq \underbrace{2\|A_{FS}\|_2^2 q_{2\mu_F}(t_{\text{end}})}_{=:\nu} \left(\gamma + \int_0^t \|\theta(t)\|_2^2 dt\right). \qquad (3.80)$$

Using (3.72), (3.79) and (3.80), we can estimate the coupled error E_{MOR}:

$$\begin{aligned}\int_0^{t_{\text{end}}} \|E_{MOR}(t)\|_2^2 dt &\leq \int_0^{t_{\text{end}}} \|E_{F,MOR}(t)\|_2^2 dt + \int_0^{t_{\text{end}}} \|\rho(t)\|_2^2 dt + \int_0^{t_{\text{end}}} \|\theta(t)\|_2^2 dt \\ &\leq \nu\gamma t_{\text{end}} + \nu t_{\text{end}} \int_0^{t_{\text{end}}} \|\theta(t)\|_2^2 dt + \gamma + \int_0^{t_{\text{end}}} \|\theta(t)\|_2^2 dt \\ &= (1+\nu t_{\text{end}}) \left(\gamma + \int_0^{t_{\text{end}}} \|\theta(t)\|_2^2 dt \right). \end{aligned} \tag{3.81}$$

The integral in (3.81) can be further estimated by (3.79) and we get an estimation that only depends on the norm of of the partitioned matrices, the integration time t_{end} and the size of the reduced order slow subsystem r. The following proposition summerises the previous results.

Lemma 8. *Let be given a linear, partitioned ODE-IVP* (3.62) *that fulfils the Assumptions 6 and 7 (weak coupling and stability). Furthermore, let be applied a balanced truncation model order reduction to the slow changing subsystem. Then the coupled system with a reduced order, slow subsystem* (3.68) *of dimension r is again stable by choosing r sufficiently large and the error that is caused by model order reduction can be bounded in time domain by*

$$\int_0^{t_{end}} \|E_{MOR}(t)\|_2^2 dt \leq (1+\nu t_{end}) \left(\gamma + \frac{\hat{\eta}}{\hat{\varphi}} (e^{\hat{\varphi} t_{end}} - 1) \right).$$

The constants ν, γ, $\hat{\eta}$, $\hat{\varphi}$ are explained in the calculation above.

An MOR error bound for an altervative coupling

For systems, where the coupling slow-to-fast is performed via the input u_F a similar bound for the MOR caused error can be derived. Such a coupling can be written as

$$\begin{pmatrix} \dot{w}_F(t) \\ \dot{w}_S(t) \end{pmatrix} = \begin{pmatrix} A_{FF} & 0 \\ A_{SF} & A_{SS} \end{pmatrix} \begin{pmatrix} w_F(t) \\ w_S(t) \end{pmatrix} + \begin{pmatrix} B_{F1} w_S(t) \tilde{u}_{F1}(t) + u_{F2}(t) \\ B_S u_S(t) \end{pmatrix}, \qquad w_0 = \begin{pmatrix} w_{F0} \\ w_{S0} \end{pmatrix} \tag{3.82}$$

with $\tilde{u}_{F1}(t) = \text{diag}(u_{F1}^{(1)}(t), \ldots, u_{F1}^{n_S}(t))$, scalar input functions $u_{F1}^{(i)} : [0, t_{\text{end}}] \to \mathbb{R}$, $i = 1, \ldots, n_S$, an input-coupling matrix $B_{F1} \in \mathbb{R} n_F \times n_S$ and $u_{F2} : [0, t_{\text{end}}] \to \mathbb{R}^{n_F}$. To compute the reduced order, slow subsystem we set again $C = I_{n_S}$ and the reduced order, coupled system reads

$$\begin{pmatrix} \dot{\tilde{w}}_F(t) \\ \dot{w}_S^r(t) \end{pmatrix} = \begin{pmatrix} A_{FF} & 0 \\ A_{SF}^r & A_{SS}^r \end{pmatrix} \begin{pmatrix} \tilde{w}_F(t) \\ w_S^r(t) \end{pmatrix} + \begin{pmatrix} B_{F1}(\mathcal{V} w_S^r(t)) \tilde{u}_{F1}(t) + u_{F2}(t) \\ B_S^r u_S(t) \end{pmatrix} \tag{3.83}$$

with the same notation as for (3.68) and initial values $w_{F,0}$ and $\mathcal{W}^\top w_{S,0}$. The error which is caused by the model order reduction error is defined analogously to (3.71)

$$\begin{aligned} \|E_{MOR}^*(t)\|_2^2 &= \left\| \begin{pmatrix} E_{F,MOR}^*(t) \\ E_{S,MOR}^*(t) \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} w_F(t) - \tilde{w}_F(t) \\ w_S(t) - \mathcal{V} w_S^r(t) \end{pmatrix} \right\|_2^2 & (3.84) \\ &\leq \left\| E_{F,MOR}^*(t) \right\|_2^2 + \|\rho^*(t)\|_2^2 + \|\theta^*(t)\|_2^2. & (3.85) \end{aligned}$$

Since the definition of the slow changing subsystem in (3.82) is not changed from the original setting in (3.62), the error terms $\rho^*(t)$ and $\theta^*(t)$ can be estimated analogously to the original setting. To derive a bound for the error of the fast subsystem, the original prove has to be modified slightly. We write $E^*_{F,MOR}(t)$ as ODE initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}t}E^*_{F,MOR}(t) = A_{FF}E^*_{F,MOR}(t) + B_{F1}\left(\rho^*(t)+\theta^*(t)\right)\tilde{u}_{F1}(t), \quad E^*_{F,MOR}(0) = 0$$

with the analytical solution

$$E^*_{F,MOR}(t) = \underbrace{\int_0^t e^{A_{FF}(t-s)}B_{F1}\rho^*(s)\tilde{u}_{F1}(s)\mathrm{d}s}_{=:\tilde{\kappa}^*(t)} + \underbrace{\int_0^t e^{A_{FF}(t-s)}B_{F1}\theta^*(s)\tilde{u}_{F1}(s)\mathrm{d}s}_{=:\tilde{\varsigma}^*(t)}.$$

We define the functions $\kappa^*(t) := \mathcal{W}^\top A_{SF}\tilde{\kappa}^*(t)$ and $\varsigma^*(t) := \mathcal{W}^\top A_{SF}\tilde{\varsigma}^*(t)$. Analogously to the derivation of the inequality (3.75) we get

$$\begin{aligned}\|\hat{\theta}(t)\|_2 \leq \|\mathcal{W}^\top A_{SS}\|_2 \int_0^t \|e^{A^r_{SS}(t-s)}\|_2\|\rho^*(s)\|_2 ds \\ + \int_0^t \|e^{A^r_{SS}(t-s)}\|_2\|\kappa^*(s)\|_2 ds + \int_0^t \|e^{A^r_{SS}(t-s)}\|_2\|\varsigma^*(s)\|_2 ds.\end{aligned} \tag{3.86}$$

We will now estimate the involved terms, starting with $\kappa^*(t)$:

$$\begin{aligned}\|\kappa^*(t)\|_2 &\leq \|\mathcal{W}^\top A_{SF}\|_2\|B_{F1}\|_2 \left(\int_0^t (e^{\mu_F(t-s)})^2\mathrm{d}s\right)^{1/2}\left(\int_0^t \|\rho^*(s)\|_2^2\|u_{F1}(s)\|_2^2\mathrm{d}s\right)^{1/2} \\ &\leq \|\mathcal{W}^\top A_{SF}\|_2\|B_{F1}\|_2 \left(\int_0^t (e^{\mu_F(t-s)})^2\mathrm{d}s\right)^{1/2}\left(\int_0^t \|\rho^*(s)\|_2^2(\max_{\tau\in[0,s]}\|u_{F1}(\tau)\|_2)^2\mathrm{d}s\right)^{1/2} \\ &\leq \underbrace{\gamma\cdot\|\mathcal{W}^\top A_{SF}\|_2\|B_{F1}\|_2 \max_{\tau\in[0,t_{\mathrm{end}}]}\|u_{F1}(\tau)\|_2\left(q_{2\mu_F}(t_{\mathrm{end}})\right)^{1/2}}_{=:\varphi^*}\end{aligned} \tag{3.87}$$

In a similar way, we derive

$$\|\varsigma^*(t)\|_2 \leq \varphi^*\left(\int_0^t \|\theta^*(s)\|_2^2\mathrm{d}s\right)^{1/2}.$$

The remaining proof of the error bound for the coupled system (3.82), is done analogously to the original proof starting in line (3.77), using the redefined constant φ^*.

Corollary 9. *Let be given a linear, partitioned ODE-IVP* (3.82) *that fulfils the Assumptions 6 and 7 (weak coupling and stability). A balanced truncation model order reduction is applied to the slow changing subsystem. The coupled system with a reduced order, slow subsystem* (3.83) *of dimension r is again stable by choosing r sufficiently large and the error that is caused by model*

order reduction can be bounded in time domain by

$$\int_0^{t_{end}} \|E^*_{MOR}(t)\|_2^2 dt \leq (1+\nu^* t_{end}) \left(\gamma + \frac{\hat{\eta}^*}{\hat{\varphi}^*} (e^{\hat{\varphi}^* t_{end}} - 1) \right)$$

with

$$\begin{aligned}
\nu^* &= 2\|B_{F1}\|_2^2 q_{2\mu_F}(t_{end})(\max_{\tau\in[0,t_{end}]} \|u_{F1}(\tau))^2 \\
\hat{\eta}^* &= 2\|\mathcal{V}\|_2^2 \gamma^2 \left(\|\mathcal{W}^\top A_{SS}\|_2 (q_{2\mu_S}(t_{end}))^{1/2} + \varphi^* q_{\mu_S}(t_{end}) \right)^2 \\
\hat{\varphi}^* &= 2\|\mathcal{V}\|_2^2 {\varphi^*}^2 q_{2\mu_S}(t_{end}) t_{end}
\end{aligned}$$

and φ^* *given in* (3.87).

After deriving bounds for the MOR caused error an estimate for E_{MR} is provided in the next section for the particular case of a balanced, slow subsystem.

3.3.4 An Error Recursion for E_{MR} for Balanced Systems

It remains to find an estimation of the multirate time integration error for the reduced order, coupled system. Usually, the time multirate integration method and its integration parameters H and m are chosen according to the properties of the reduced order, coupled system. For this setting, an estimate for E_{MR} can be easily found by using the the results of the convergence analysis for multirate method of Section 2.2.

Now, we assume to have a stable multirate time integration method for the full order, coupled system (3.62). We study the question, how the size of the reduced order, slow subsystem r has to be chosen, such that the multirate time integration is still stable. For the integration scheme, we choose a multirate θ-method with a Coupled-Slowest-First approach [HS09]. The integration method for the macro-step $t_n \to t_{n+1} = t_n + H$ reads

$$\begin{aligned}
\begin{pmatrix} w^*_{F,n+1} \\ w_{S,n+1} \end{pmatrix} = \begin{pmatrix} w^*_{F,n} \\ w_{S,n} \end{pmatrix} &+ \theta H \left(\begin{pmatrix} A_{FF} & A_{FS} \\ A_{SF} & A_{SS} \end{pmatrix} \begin{pmatrix} w^*_{F,n} \\ w_{S,n} \end{pmatrix} + \begin{pmatrix} u_F(t_n) \\ B_S u_S(t_n) \end{pmatrix} \right) \\
&+ (1-\theta) H \left(\begin{pmatrix} A_{FF} & A_{FS} \\ A_{SF} & A_{SS} \end{pmatrix} \begin{pmatrix} w^*_{F,n+1} \\ w_{S,n+1} \end{pmatrix} + \begin{pmatrix} u_F(t_{n+1}) \\ B_S u_S(t_{n+1}) \end{pmatrix} \right)
\end{aligned} \tag{3.88}$$

for a parameter $\theta \in [0,1]$. The approximation $w^*_{F,n+1}$ is inaccurate and therefore refused and only $w_{S,n+1}$ is used as proper approximation. For the micro-steps we assume a fixed multirate factor $m = 2$ and the integration of the fast subsystem is given by

$$\begin{aligned}
w_{F,n+(i+1)/2} = w_{F,n+i/2} &+ \theta \tfrac{H}{2} \left(A_{FF} w_{F,n+i/2} + A_{FS} \bar{w}_{S,n+i/2} + u_F(t_{n+i/2}) \right) \\
&+ (1-\theta) \tfrac{H}{2} \left(A_{FF} w_{F,n+(i+1)/2} + A_{FS} \bar{w}_{S,n+(i+1)/2} + u_F(t_{n+(i+1)/2}) \right)
\end{aligned} \tag{3.89}$$

for $i = 0, 1$, with $t_{n+i/2} = t_n + \frac{iH}{2}$ and $\bar{w}_{S,n+1/2}$ the interpolated value of w_S at $t_{n+1/2}$. We consider the error after $n+1$ macro-steps

$$E^{MR}_{n+1} = w(t_{n+1}) - w_{n+1}$$

for $w = [w_F^\top,\ w_S^\top]^\top$ between the analytical solution $w(t_{n+1})$ of (3.62) and the numerical approximation w_{n+1} achieved by the multirate θ-method (3.88-3.89). In terms of [HS09], this error can be expressed by the error at the previous macro-step E_n^*, an amplification matrix S and the local truncation error z_{n+1}

$$E_{n+1}^{MR} = SE_n^{MR} + z_{n+1}. \tag{3.90}$$

In [HS09] it is shown that the order of z_n is 2 for $\theta = 0.5$ and 1 otherwise. We assume that the multirate θ-method for macro-step size H and multirate factor 2 applied to the full order, coupled systen (3.62) is stable, i.e.,

$$\|S^n\|_\infty < D \quad \forall\, n \geq 0. \tag{3.91}$$

for a constant $D > 0$. Now, we show that the stability of the multirate θ-method for the full order, coupled system (3.62) can be used to derive a stability condition for the multirate θ-method applied to the reduced order, coupled system (3.68). To this end, we recall a perturbation condition which is given in [HS09]: Let (3.91) hold for (3.62) and denote

$$A = \begin{pmatrix} A_{FF} & A_{FS} \\ A_{SF} & A_{SS} \end{pmatrix}$$

the system matrix of the full order, coupled system. Let $\tilde{A}$ be the matrix of a perturbed system such that $\|A - \tilde{A}\|_\infty \leq L$ for a moderate constant L. Then the amplification matrix $\tilde{S}$ of the perturbed system can be bounded by

$$\|(\tilde{S})^n\|_\infty \leq De^{dDt_{\text{end}}} \tag{3.92}$$

with a constant $d = d(L,D)$. In our setting, the perturbed matrix $\tilde{A}$ is given by the system matrix of the reduced order, coupled system (3.68), identified in the original vector space

$$\tilde{A} = \begin{pmatrix} A_{FF} & A_{FS}\mathcal{V}\mathcal{W}^\top \\ \mathcal{V}\mathcal{W}^\top A_{SF} & \mathcal{V}\mathcal{W}^\top A_{SS}\mathcal{V}\mathcal{W}^\top \end{pmatrix}. \tag{3.93}$$

For the particular case of a balanced, slow subsystem, we will derive the system matrix $\tilde{A}$ and measure the perturbation L. For a balanced system, the projection matrices for balanced truncation are given by $\mathcal{V} = \mathcal{W} = (I,\ 0)^\top$, cf. [Ant05]. Let the system matrices of the slow subsystem in (3.62) be of the following form

$$A_{SS} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad A_{SF} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \qquad A_{FS} = (C_1,\ C_2),$$

where A_{11}, B_1 and C_1 describe the states of the system that will kept by the model order reduction. Then the system matrix of the coupled system with reduced order, slow subsystem reads

$$\begin{pmatrix} A_{FF} & C_1 \\ B_1 & A_{11} \end{pmatrix}.$$

Since the system properties of the reduced system approximate the properties of the original system the influence of the truncated part on the kept part can be assumed to be small. We formalise this in the following Assumption.

Assumption 10. *We assume that*

$$\|A_{12}\|_\infty \leq \|A_{22}\|_\infty \tag{3.94}$$

holds.

The system matrix $\tilde{A}$ of the perturbed system is given by

$$\tilde{A} = \begin{pmatrix} A_{FF} & C_1 \mathcal{W}^\top \\ \mathcal{V} B_1 & \mathcal{V} A_{11} \mathcal{W}^\top \end{pmatrix} = \begin{pmatrix} A_{FF} & C_1 & 0 \\ B_1 & A_{11} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.95}$$

and we find the perturbation

$$L = \|A - \tilde{A}\|_\infty = \left\| \begin{pmatrix} 0 & 0 & C_2 \\ 0 & 0 & A_{12} \\ B_2 & A_{21} & A_{22} \end{pmatrix} \right\|_\infty .$$

Using Assumptions 6, 7 and 10 yields

$$\|A - \tilde{A}\|_\infty \leq 2\varepsilon + 3\|A_{22}\|_\infty =: \tilde{\varepsilon}.$$

The previous results are summarised in the following lemma:

Lemma 11. *Let be given system* (3.62) *which fulfils the Assumptions 6, 7 and 10 and its slow subsystem is balanced. Let the multirate θ-method be stable and the reduced order, slow subsystem be achieved by balanced truncation. Then the multirate θ-method is also stable for the coupled system with reduced order, slow subsystem. The amplification matrix $\tilde{S}$ for the error is bounded by*

$$\|(\tilde{S})^m\|_\infty \leq M e^{\tilde{D} D t_{end}}$$

with $\tilde{D} = \tilde{D}(D, \tilde{\varepsilon})$.

This results describe the influence of the model order reduction applied to the slow subsystem on the stability of the coupled system, on the stability of the multirate θ-method and on the split error in (3.70). For this special case of a balanced, slow subsystem the it can be directly seen that the dimension of the reduced subsystem r impacts the size of the perturbation L and therefore also the time integration error E_{MR}.

Considering the estimates for the error that is caused by the model order reduction E_{MOR} and the time integration error E_{MR} one notices that the estimates ate given in different norms. The following theorem summarises both error bounds and adapts the norm of both errors.

Proposition 12. *Let be given a linear, partitioned ODE system 3.62 on the time interval $[0, t_{end}]$ that fulfils the Assumptions* (3.63)*,* (3.64) *and 10. Let the slow subsystem be balanced and let be applied a balanced truncation model order reduction to the slow subsystem. A multirate θ-method is used to integrate the partitioned ODE system with macro-steps t_{end}/m and micro-steps $t_{end}/2m$. Then the error between the analytical solution of the original, non-reduced system and the multirate θ-method approximation of the system with a reduced order, slow subsystem reads*

$$\int_0^{t_{end}} \|E(t)\|_2^2 dt \leq (1 - \nu t_{end}) \left(\gamma + \frac{\hat{\eta}}{\hat{\varphi}} \left(e^{\hat{\varphi} t_{end}} - 1 \right) \right) + (n_F + n_S) D^2 t_{end} \left(\sum_{i=1}^{n} e^{d D t_i} \|z_i\|_2 \right)^2$$

with the above notation for ν, γ, $\hat{\eta}$, $\hat{\varphi}$, d, C and z_i, macro-steps $t_i = i \cdot t_{end}/n$.

We have $\|z_i\|_2 = \mathcal{O}(H^{p+1})$ with $p = 2$ in case of $\theta = 0.5$ and $p = 1$ otherwise.

The factor $(n_F + n_S)$ in the second summand is caused by the equivalence between the 2- and the ∞-norm. Although both norms are equivalent the equivalence constant depends on the dimension of the full order system and is therefore quite large. We point out that the coupled error estimate is more a theoretical aspect of the analysis than a practical calculation.

Section Summary

We investigated the influence of a model order reduction applied to the slow changing subsystem of a linear, multiscale partitioned ODE-IVP. We derived an error bound in time domain for the coupled ODE-IVP with order reduced, slow subsystem. For the multirate θ-method, we provided a combined error bound that estimates the influence of the model order reductions as well as the time integration error.

In some applications, the dynamical behaviour of the real-world systems cannot be described by linear-affine systems. In that case, we refer to the proceeding work of [BCG20]. Here, the authors consider and investigate model order reduction techniques for non-linear dynamical systems in combination with multirate time integration.

The following, last section of this chapter leads over to the subsequent Chapter 4 about multirate time integration for differential-algebraic equation. We present a model order reduction technique and a multirate time integration method for a field-circuit coupled system, where the mathematical model of the electromagnetic field leads to a differential-algebraic equation.

3.4 Multirate Time Integration and Model Order Reduction for a Field-Circuit Coupled System

Results of this section have been published partially in [HBG+18].

Often, the mathematical modelling of a technical or physical system does not lead to a system of ordinary differential equations. For a space dependent partial differential equation (PDE) a semi-discretisation of the spatial variable leads back to a systems of ODEs as we saw in Section 3.2.4 using the example of the heat equation. In other cases, additional algebraic constraints arise that have to be fulfilled by the solution of the differential equation. For the numerical treatment of such differential-algebraic equations (DAEs) the classical ODE setting has to be modified.

In the following, we consider a field-circuit coupled system where the circuit is assumed to provide high dynamical changes. Due to the physical properties and sizes the electromagnetic field is changing much slower. We start with the modelling of electromagnetic field which is described by a PDE and its semi-discretisation leads to a system of DAEs. The multirate methods presented in Section 2.2 have to be adapted for coupled systems of DAEs. Multirate methods for DAEs have been first discussed in [Str06] and [Ver08]. In this chapter, we only present a rough idea about multirate schemes for linear DAEs and will postpone a detailed and more general analysis for multirate time integration for DAEs to Chapter 4. Since the resulting slow changing field system is of large dimension, we apply a model order reduction to the subsystem. We use the method of [KBS17] that projects the large-scale DAEs to a small dimensional system of ODEs. Simulation results for the non-reduced coupled system of DAEs and the reduced order system of ODE are given at the end of this section.

Mathematical Modelling

Our multi-physics benchmark system consists of an electric circuit and the electromagnetic field of a single-phase 2D-transformer with an iron core and two coils. Figure 3.5 shows a circuit diagram of the coupled system, where the electromagnetic effects are represented by the lumped devices of a transformer in the box.

The mathematical models of the electric circuit and the electromagnetic field of the transformer are coupled by the source coupling approach [BBGS13]. That is, add an additional controlled current source to the circuit subsystem which is realised by the output of the field subsystem $i_M(t)$. And we add an additional voltage source to the transformer's subsystem which is given by the solution of the circuit subsystem $u_1(t)$. The procedure is illustrated in Figure 3.6. We start with the modelling of the single subsystems.

Circuit Modelling. The circuit of our benchmark system given in Figure 3.6 is described by one ODE for the node potential u_1

$$C\tfrac{d}{dt}u_1(t) = G(u_1(t) - U_{in}(t)) - i_M(t) \tag{3.96}$$

with capacitance $C = 1\text{nF}$, conductance $G = 10^{-3}S$, the input voltage $U_{in}(t) = 45.5 \cdot 10^3 \sin(900\pi t) + 10^3 \sin(45000\pi t)$ and the (later specified) coupling term $i_M(t)$. The here used variable G refers to the physical quantity of conductance and is independent of the notation of the transfer function in (3.24). The presented multirate method can also be applied to more general electrical circuits which are usually described by a system of DAEs [GF99].

Field Modelling. For the electromagnetic field subsystem, we consider a magneto-quasistatic (MQS) problem which is described by Maxwell's equation in the magnetic vector potential formulation

$$\begin{aligned} \sigma\frac{\partial \hat{A}}{\partial t} + \nabla \times \left(\nu \nabla \times \hat{A}\right) &= J && \text{in } \Omega \times (0, t_{\text{end}}) \\ \text{with boundary conditions} \quad \hat{A} \times n_0 &= 0 && \text{on } \partial\Omega \times (0, t_{\text{end}}) \\ \text{and initial conditions} \quad \hat{A} &= \hat{A}_0 && \text{in } \Omega. \end{aligned} \tag{3.97}$$

$\Omega = \Omega_C \cup \Omega_N$ is a bounded two-dimensional domain composed of a conducting Ω_C and non-conducting subdomain Ω_N. $\hat{A}$ is the magnetic vector potential, $\nu = \nu(\hat{A})$ is the magnetic reluctivity with $\nu_C = 14872\text{Am/(Vs)} = 14872\text{m/H}$ on Ω_C and $\nu_N = 1\text{Am/(Vs)} = 1\text{m/H}$ on Ω_N. $\sigma = 5 \cdot 10^5 \Omega^{-1}\text{m}^{-1}$ is the electric conductivity which vanishes on Ω_N. n_0 is the outer unit normal vector to the boundary $\partial\Omega$ of Ω. J denotes the current density applied by external sources. Taking $J = \chi i_M$ with a divergence-free winding function χ and i_M the vector of lumped currents through the transformer, then the coupling term to the circuit subsystem can be written as

$$\int_\Omega \chi^\top \tfrac{\partial}{\partial t}\hat{A}\, d\xi + R i_M = u_1, \tag{3.98}$$

where R is the resistance matrix and u_1 describes the applied voltage by the circuit to the transformer. Applying the finite element discretisation method to (3.97) and (3.98) and reordering unknown variables accordingly to the conducting and non-conducting subdomains, we can obtain

a linear system of DAEs

$$\mathfrak{M}\tfrac{d}{dt}\begin{bmatrix} a \\ i_M \end{bmatrix} = \mathfrak{F}\begin{bmatrix} a \\ i_M \end{bmatrix} + \mathfrak{B}u, \qquad y_S = i_M = \mathfrak{C}\begin{bmatrix} a \\ i_M \end{bmatrix} \tag{3.99}$$

with a singular mass matrix $\mathfrak{M}$, a semi-discretised vector of magnetic potentials a, an input u_1 which is the voltage at the primary coil and the output $y_S = i_M$ the current through the primary coil. For this particular example, we have $\mathfrak{C} = \mathfrak{B}^\top = [0, \ldots, 0, 1]$. The FEM discretisation is done by the free available software FEniCS[1] for the dimension $n_S = 7823$.

Model Order Reduction for Magneto-Quasistatic Equation

We briefly discuss model order reduction of the MQS equations (3.99), for more details, we refer to [KBS17].

The properties of the involved system matrices guarantee that the DAE system (3.99) can be transformed into a system of ODEs

$$M\dot{w}_S = \bar{F}w_S + Bu_1, \qquad y_S = i_M = -B^\top M^{-1}\bar{F}w_S \tag{3.100}$$

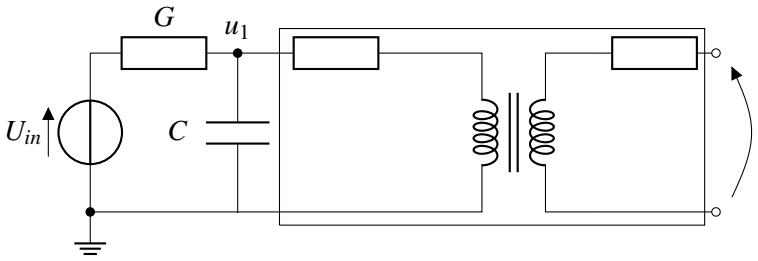
with nonsingular, symmetric, positive definite matrices M and $-\bar{F}$ and a corresponding vector of unknowns $w_S = w_S(t)$. This transformation is possible since the magnetic reluctivity ν is assumed to be constant on Ω_C and the DAE system (3.99) is of index 1 [KBS17]. A detailed definition of the index of a DAE is given in Section 4.1.1. Note that system (3.100) has the same input u_1 and the same output i_M as the DAE system (3.99) meaning that the input-output relation of (3.99) is preserved in (3.100).

System (3.100) is asymptotically stable and passive [KBS17]. For model order reduction of (3.100), we use a balanced truncation approach as described in Section 3.1.2 based on the reachability Gramian P which is defined as a unique symmetric and positive semidefinite solution to the generalized Lyapunov equation

$$\bar{F}PM + MP\bar{F} = -BB^\top. \tag{3.101}$$

Due to the symmetry conditions, the observability Gramian Q satisfies $MQM = \bar{F}P\bar{F}$. Let $P = SS^\top$

Figure 3.5: Circuit diagram of the coupled systems with lumped elements for the electromagnetic effects (box).



[1] http://fenicsproject.org

be a Cholesky factorization of P. We compute the eigenvalue decomposition

$$-S^\top \bar{F} S = [U_1,\, U_0]\,\mathrm{diag}(\Lambda_1,\, \Lambda_0)\,[U_1,\, U_0]^\top,$$

where Λ_1 and Λ_0 are diagonal matrices and Λ_1 contains all kept Hankel singular values and Λ_0 all truncated ones. Now, we can determine the reduced-order model by projection

$$M^r \dot{w}_S^r = \bar{F}^r w_S^r + B^r u, \qquad \tilde{y}_S = C^r w_S^r, \tag{3.102}$$

where $M^r = \mathcal{W}^\top M \mathcal{V}$, $\bar{F}^r = \mathcal{W}^\top \bar{F} \mathcal{V}$, $B^r = \mathcal{W}^\top B$ and $C^r = -B^\top M^{-1} \bar{F} \mathcal{V}$ with the projection matrices $\mathcal{V} = S U_1 \Lambda_1^{-1/2}$ and $\mathcal{W} = -M^{-1} \bar{F} \mathcal{V}$. One can show that the reduced matrices M^r and $-\bar{F}^r$ are symmetric, positive definite and $C^r = B^{r\top}$ guarantees that system (3.102) is passive. Moreover, we have the L_2-norm error bound for the output

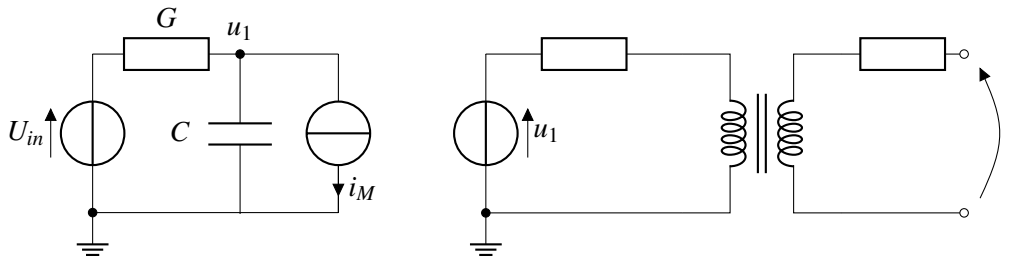
$$\|y_S - \tilde{y}_S\|_2 \leq 2\,\mathrm{trace}(\Lambda_0)\,\|u\|_2.$$

For solving the generalized Lyapunov equation (3.101), we can use the low-rank alternating direction implicit method or (rational) Krylov subspace method [BS13, Sim16]. In both methods, we need to solve linear systems of the form $(\tau M + \bar{F})v = b$ for a vector v with possibly dense M and $\bar{F}$. Both, M and $\bar{F}$ provide a block structure which can be exploited for the construction of the linear system of equations. Doing so, a connection to the full order matrices $\mathfrak{M}$ and $\mathfrak{F}$ of the DAE system (3.99) can be derived and an equivalent system of linear equations $(\tau\mathfrak{M} + \mathfrak{F})\hat{v} = \hat{b}$ with the sparse matrices $\mathfrak{M}$ and $\mathfrak{F}$ can be constructed.

The order reduced, coupled system (3.96 & 3.102) yields a system of ODEs and can be simulated in time domain efficiently with a multirate time integration method as presented in Section 2.2.

For the non-reduced, coupled system (3.96 & 3.99), we end up with a linear DAE system and multirate methods have to be derived for this class of problems. For linear DAEs of index-1, time integration methods for linear-implicit ODEs can be applied. In the following section, we will deduce a multirate method for linear DAEs from the LobattoIIIC scheme.

Figure 3.6: Source coupling approach: Circuit diagram of the subsystem with additional current source i_M in the circuit subsystem and an additional voltage source u_1 in the field subsystem.



Multirate Time Integration for Linear ODE/DAE-Systems

We consider the non-reduced, coupled ODE/DAE-system (3.96 & 3.99), set $x = [a^\top,\ i_M^\top]^\top$ and get the coupled DAE initial value problem

$$\underbrace{\begin{pmatrix} C & 0 \\ 0 & \mathfrak{M} \end{pmatrix}}_{=:\check{M}} \begin{pmatrix} u_1 \\ \dot{x} \end{pmatrix} = \underbrace{\begin{pmatrix} G & -\mathfrak{B}^\top \\ \mathfrak{B} & \mathfrak{F} \end{pmatrix}}_{=:A} \underbrace{\begin{pmatrix} u_1 \\ x \end{pmatrix}}_{=:\check{x}} - \underbrace{\begin{pmatrix} G \\ 0 \end{pmatrix}}_{=:\check{B}} U_{in}(t), \qquad x(t_0) = x_0,\ u_1(t_0) = u_{10} \tag{3.103}$$

with a singular mass-matrix diag $\check{M}$. The field subsystem (3.99) and the coupled DAE system (3.103) are of index-1 [KBS17, HBG+18]. Such DAE-systems can be integrated with an implicit Runge-Kutta method [BCP95]. To derive a multirate Runge-Kutta method for linear DAEs of index-1, we use the LobattoIIIC scheme, its Butcher tableau is given by

$$\begin{array}{c|cc} 0 & \frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}.$$

For the multirate scheme, we use the Coupled-Slowest-First approach: We integrate the coupled DAE system (3.103) on the macro-step grid and consider the time step $t_n \to t_n + H$. To compute the increments $k_1,\ k_2$, the linear DAE system (3.103) leads to the following linear system of equations

$$\begin{pmatrix} \check{M} - \frac{H}{2}A & \frac{H}{2}A \\ -\frac{H}{2}A & \check{M} - \frac{H}{2}A \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} A\check{x}_n - \check{B}U_{in}(t_n) \\ A\check{x}_n - \check{B}U_{in}(t_n + H) \end{pmatrix}$$

Then, the approximation at $t_n + H$ is given by

$$\check{x}_{n+1} = \begin{pmatrix} \tilde{u}_{1n+1} \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} u_{1n} \\ x_n \end{pmatrix} + \frac{H}{2}(k_1 + k_2).$$

Since k_1 and k_2 are derived by solving a *linear* system of eqautions, only linear effects of the approximation properties of the LobattoIIIC scheme are illustrated in the following.

The macro-step size H is chosen according to the properties of the slow changing, field subsystem. Therefore, the approximation x_{n+1} is accepted, the approximation of the circuit subsystem $\tilde{e}_{n+1}$ is not sufficiently accurate and therefore refused.

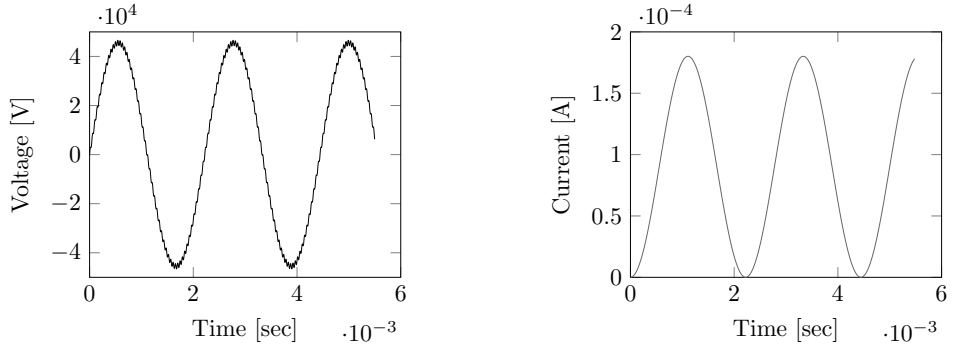
To achieve an appropriate approximation for u_{1n+1}, m integration steps of micro-step size h are carried out. For the micro-step $(t_n + lh) \to (t_n + (l+1)h)$ with $l = 0, \ldots, m-1$, the increments $k_{1,l}^F,\ k_{2,l}^F$ are given by the following system of linear equations

$$\begin{pmatrix} C - \frac{h}{2}G & \frac{h}{2}G \\ -\frac{h}{2}G & C - \frac{h}{2}G \end{pmatrix} \begin{pmatrix} k_{1,l}^F \\ k_{2,l}^F \end{pmatrix} = \begin{pmatrix} Gu_{1n+lh} - \mathfrak{B}^\top \bar{x}_{n+lh} - U_{in}(t_n + lh) \\ Gu_{1n+lh} - \mathfrak{B}^\top \bar{x}_{n+(l+1)h} - U_{in}(t_n + (l+1)h) \end{pmatrix}.$$

with interpolated values $\bar{x}_{n+lh}$ at $t_n + lh$. We get the intermediate approximations by

$$u_{1n+(l+1)h} = u_{1n+lh} + \frac{h}{2}\left(k_{1,l}^F + k_{2,l}^F\right).$$

The derived multirate time integration scheme for linear DAEs can be also applied to implicit ODE-systems like the reduced order, coupled system (3.96 & 3.102). The simulation results for both coupled system are summarised in the following section.



(a) Node Potential of e_1

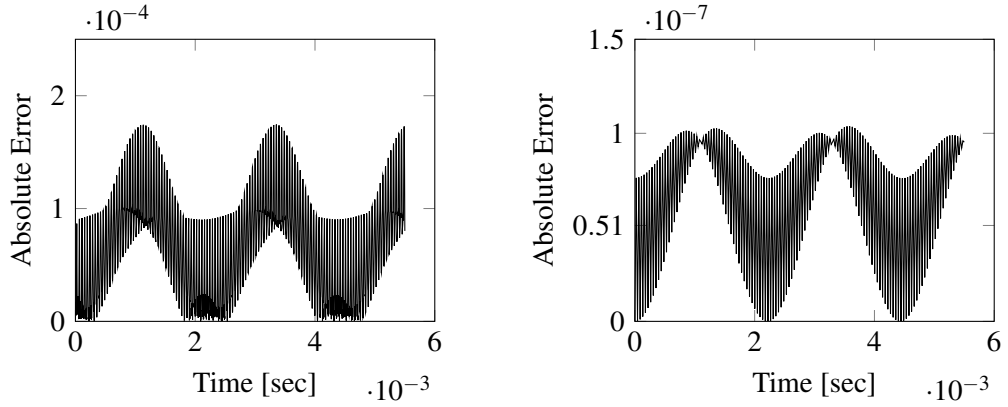
(b) Current through the Primary Coil

Figure 3.7: Numerical solution of the subsystems.

Simulation Results

We integrate the coupled DAE system by the multirate LobattoIIIC scheme over the time interval [0s, 0.0055s]. Since we are interested in the influence of the multirate approach, we consider a reference solution that is computed by the single-rate LobattoIIIC method with constant global step size using 2500 time steps. We also integrated the coupled system with constant global step size using the double amount of time steps. The maximum relative 2-norm error in the outputs of the subsystems between both solutions was $3.9 \cdot 10^{-3}$. We accepted the 2500 time step solution as reference solution with a moderate accuracy. The simulation was run on a Intel Core2 Duo P7450 with 2.13GHz with 4GB RAM. For the coupled full-order, coupled DAE system the computation time was 728.2s. Figure 3.7 shows the outputs of the two subsystems: (a) the node potentials u_1, which belongs to the fast changing subsystem (basically we see the superposition of the sinusoidal oscillations) and (b) the current through the primary coil of the transformer, which belongs to the slow subsystem.

To investigate the influence of the multirate approach on the full order DAE system, the time interval is discretised into 250 macro-step and each macro-step is refined into 10 micro-steps. 250 macro-steps are sufficient to integrate the slow changing, field subsystem and 2500 micro-steps are needed for the fast changing, circuit subsystem to reach an adequate approximation. Here, the computation ended after 77.4s. We computed the error between the single-rate and the multirate approximation separately for both subsystems. For the fast changing subsystem, the error is computed by the absolute value of the difference between the node potential of the single-rate solution and the node potential achieved by the multirate approximation at each micro-step. In Section 4.3 we show, that the approximation of the fast, differential variable on the micro-step grid is as accurate as the approximation of the slow subsystem on the macro-step. Therefore, we investigate the error of the fast subsystem on the micro-step grid. For the slow changing subsystem, we computed the absolute value of the difference in the output of the subsystem i_M on the macro-step grid. Figure 3.8 illustrates these errors. In the fast changing subsystem the error increases during one macro-step since there is an additional error that is caused by interpolating the values of the slow changing subsystem. At the macro-steps the subsystems are integrated together, so that the error at these time points is usually smaller. In the slow subsystem, every second approximation gives better results while the intermediate approximation is worse. Until now, this phenomena is not yet understood completely. Since the size of the error is in total small, the improvement in computation time motivates and justifies the usage of multirate time integration schemes for these DAEs.



(a) Circuit Subsystem (b) Electromagnetic Subsystem

Figure 3.8: Absolute errors in the subsystems between the single-rate reference approximation of the full-order DAE-system and the multirate approximation of the full order DAE-system.

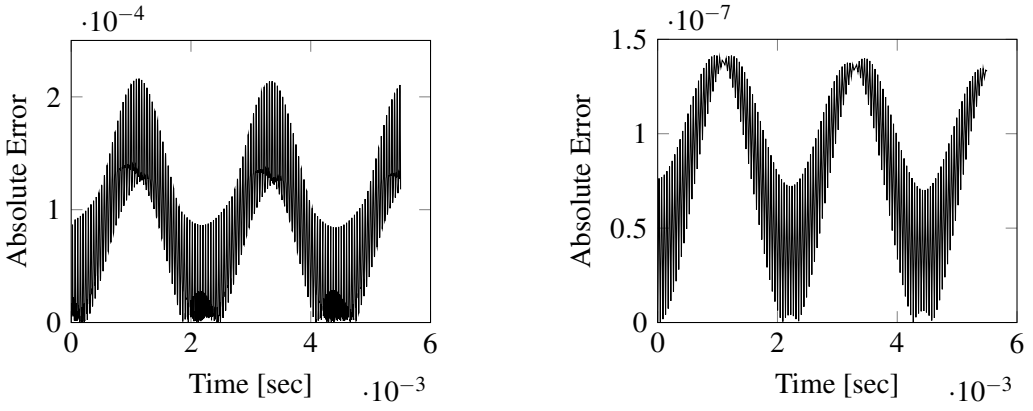
The slow changing, field subsystem is reduced to dimension $\dim(w_S^r) = 4$. The reduced-order, coupled system of ODEs is integrated by the same multirate method with the same integration parameters as for the full-order system of DAEs. The simulation needed 0.20s to compute. Figure 3.9 shows the absolute error between the single-rate reference approximation of the full-order DAE system and the multirate approximation of the coupled system of ODEs with order reduced, slow subsystem. The error for both subsystems here is very small and only slightly larger than with full-order, slow subsystem.

Finally, we integrated the coupled system with the reduced order, slow subsystem (3.102) without multirating, so we used the same integration parameters as for the DAE single-rate solution. Figure 3.10 shows the error for both subsystem which is in both cases quite small. The computation time was 0.13s, so it was a bit faster than with multirating. This phenomenon can be explained by the ratio between the number of fast and slow changing variables. In our case, the full-order system has a ratio of $1 : 7821$, while for the reduced-order system, it is $1 : 4$.

This ratio is an indicator for the gain of efficiency between the single-rate and multirate approximation. If there is a large number of slow changing variables compared to a small number of fast changing variables, a multirate time integration scheme saves many function evaluation of the large dimensional slow subsystem. However, the implementation of a multirate scheme is more complex than for a classical singlerate scheme. So if the dimension of the slow changing subsystem is only a little bit larger than the dimension of the fast changing subsystem, a multirate scheme can be even less efficient than the corresponding single-rate scheme.

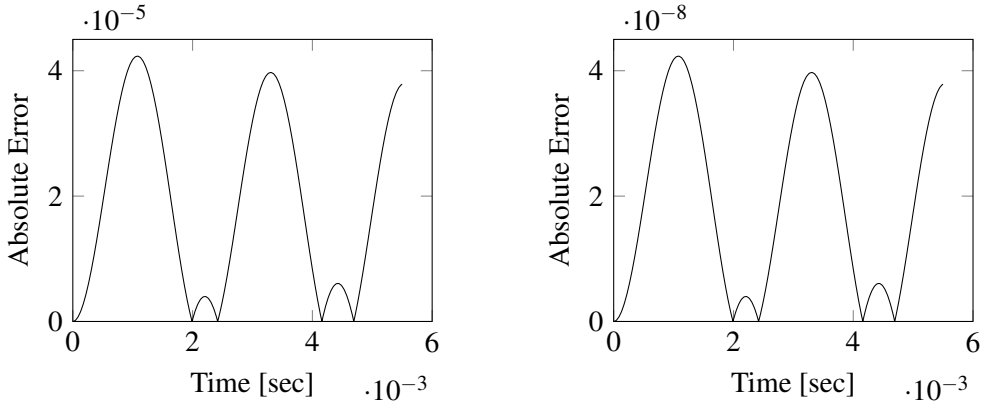
Chapter Summary

We applied a projection based, model order reduction to the slow changing subsystem of a multiscale partitioned ODE-IVP and integrated the resulting coupled system with a multirate method. The model order reduction leads to a much smaller set of slow changing variables. By introducing the interface reduction approach, we showed that beside the MOR for the slow subsystem, the choice of a small dimensional coupling interface between the subsystem is necessary to gain efficiency during the multirate time integration. Simulation results of a thermal-electric coupled



(a) Circuit Subsystem (b) Electromagnetic Subsystem

Figure 3.9: Absolute errors in the subsystem between the single-rate reference approximation of the full-order DAE-system and the multirate approximation of the reduced order ODE-system.



(a) Circuit Subsystem (b) Electromagnetic Subsystem

Figure 3.10: Absolute errors in the subsystems between the single-rate reference approximation of the full-order DAE-system and the single-rate approximation of the reduced order ODE-system.

system confirmed the better performance after applying the MOR to the slow, thermal subsystem and the interface reduction to coupling variables. To investigate the influence of an order reduced, slow subsystem on the propertied of the multirate time integration method, we derived an error bound in time domain which estimates both, the MOR caused error and the integration error of the multirate method. At the end of the chapter, we considered a field-circuit coupled system, applied an MOR to the electromagnetic field system and integrated the coupled system with a multirate method. The field subsystem is described by a system of DAEs. A particular MOR technique projects the full order system of DAEs on a reduced order system of ODEs. For both, the system of DAEs and the system of ODEs we derived a multirate method for an efficient time integration. Since DAEs arise in many applications, we will study multirate time integration methods for DAE-IVPs more detailed in the following chapter.

Chapter 4
4 Multirate One-Step Methods for Differential-Algebraic Equations

In Section 3.4, we saw that it is not possible to describe the dynamical behaviour of various technical or physical systems with ordinary differential equations (ODEs), since additional algebraic constraints arise during the mathematical modelling. Beside the semi-discretised magneto-quasistatic equation of Section 3.4, which describes the electro-magnetic field of a transformer, a differential-algebraic equation (DAE) results for example from a modified nodal analysis to simulate currents and voltages in an electrical circuit [GF99]. The mathematical model of a mechanical multibody-system also leads to system of DAEs [ESF98]. The analytical and numerical treatment of DAEs is much more challenging compared to ODEs [BCP95, HW02, KM06]. For an efficient time integration of coupled systems of DAEs with different dynamical behaviour, we derive and analyse multirate Runge-Kutta schemes for DAEs in this chapter. Multirate methods for DAEs have been introduced in [BGK02] for an efficient simulation of electrical circuits. In [Str06], mixed-multirate Runge-Kutta schemes were presented and corresponding order conditions derived. Multirate multistep methods are presented and analysed in [Ver08] using a particular stability definition.

To derive a multirate integration scheme for coupled DAEs, we adapt a classical, single-rate Runge-Kutta method which can be used for time integration of DAEs. This method is equipped with different, inherent time steps according to the dynamical properties of the subsystems. The coupling between the subsystems is realised by one of the coupling strategies of Section 2.2.2. Then, we investigate the approximation properties (consistency, convergence) of the resulting multirate Runge-Kutta method for DAEs. The chapter is organised as follows:

In Section 4.1, we start with a formal definition of a DAE and explain the index of DAEs, one of the most important concepts in the theory of DAEs. For the time integration of DAEs, we derive implicit Runge-Kutta methods. In the following sections, we extend these methods to multirate Runge-Kutta methods for semi-explicit DAEs and show convergence order 1 for a method based on the implicit Euler-scheme (Section 4.2) and convergence order 2 for a LobattoIIIC-based multirate method (Section 4.3). At the end of the chapter in Section 4.4, we derive a convergence theorem for multirate one-step methods for DAEs using the Decoupled-Slowest-First approach. This theorem links the theories of multirate time integration and dynamic iteration schemes.

4.1 An Introduction to Differential-Algebraic Equations

To derive multirate methods for DAEs, we start with the necessary theory of DAEs and their numerical treatment. In Section 4.1.1, we give the definition and several formulations for DAEs and all necessary assumptions. Subsequently, we introduce the index of a DAE, which is one of the most important concepts in the theory of DAEs and describes – roughly speaking – the differences in the analytical and numerical treatment between the considered DAE and an ODE. For

the time integration of semi-explicit DAEs of index-1, we derive implicit Runge-Kutta methods in Section 4.1.2. These methods will be extended to multirate Runge-Kutta methods for DAEs later in this chapter.

4.1.1 Differential-Algebraic Equations – Definition and Index Concept

There exist different formulations of differential-algebraic equations. We start with the most general approach and give the following

Definition 7 (Differential-Algebraic Equation (DAE))**.** *An equation on a time interval $I = [t_0, t_{end}]$ between a function $x: I \to \mathbb{R}^n$, $t \mapsto x(t)$ and its derivative $\dot{x}(t) = \frac{d}{dt}x(t)$ that is described by the root of a function*

$$F(t, x(t), \dot{x}(t)) = 0, \qquad t \in I \tag{4.1}$$

is called differential-algebraic equation (DAE), if $F: I \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable with respect to $\dot{x}$ and the rank of the Jacobian is constant with

$$rank\left(\frac{\partial F}{\partial \dot{x}}(t, x, \dot{x})\right) < n.$$

Usually, the DAE is specified with initial conditions $x(t_0) = x_0$. In the following, we refer to a DAE (4.1) with initial conditions as DAE Initial Value Problem *DAE-IVP*.

A linear DAE is defined as

$$M(t)\dot{x}(t) - A(t)x(t) - f(t) = 0 \tag{4.2}$$

for continous $M, A: I \to \mathbb{R}^{n \times n}$ and $f: I \to \mathbb{R}^n$ if $\det(M) = 0$ holds. In the case of constant coefficients, i.e. $M(t) = M$, $A(t) = A$ for all $t \in I$, we end up with a DAE of the form of (3.103), which describes the the field-circuit coupled system of Section 3.4.

An important classification for DAEs is the so-called *index* of a DAE. In [Meh15] it is described as a 'measure of difficulty in the analytical or numerical treatment of the DAE'. In the last decades, different index-concepts have been proposed in Literature. In this work, we solely refer to the differentiation or differential index [CG95].

Definition 8 (Differentiation Index)**.** *Let F in* (4.1) *be sufficiently smooth with respect to t. We consider the system of derivatives*

$$\begin{aligned} 0 &= F(t, x(t), \dot{x}(t)) \\ 0 &= \tfrac{d}{dt}F(t, x(t), \dot{x}(t)) \\ &\vdots \\ 0 &= \tfrac{d^i}{dt^i}F(t, x(t), \dot{x}(t)) \end{aligned} \tag{4.3}$$

for a given integer $i \in \mathbb{N}$. If i is the smallest number, such that a system of ODEs

$$\dot{x} = \varphi(t, x(t))$$

can be derived from (4.3) *only by algebraic transformations, then i is called the differentiation index of the DAE* (4.1)*.*

A compact overview of other established index concepts is given in [Meh15]. For linear DAEs with constant coefficients, a widely used index concept is the *Kronecker*- or *Niloptency*-index, [HW02, Ch.VII.1, Th.1.1.]. For nonlinear DAEs, other relevant index concepts are the *Perturbation*-index [HW02, Ch.VII.1, Def.1.3.], the *Tractability*-index [GM86,Mär02] and the *Strangeness*-index [KM06]. All concepts have in common, that the index of a DAE is always a natural number starting at 0 or 1. The higher the index, the more challenging is the numerical and analytical treatment of the DAE.

In many applications, a DAE model provides additional properties or structure that can be exploited to facilitate the analytical or numerical treatment. In the following, we focus on *semi-explicit* DAEs.

Definition 9 (Semi-Explicit DAE)**.** *A differential-algebraic equation of the form*

$$\dot{w}(t) = f(w(t), z(t)) \tag{4.4}$$
$$0 = g(w(t), z(t)) \tag{4.5}$$

with $t \in I = [t_0, t_{end}]$, differential variables $w: I \to \mathbb{R}^{n_w}$ and algebraic variables $z: I \to \mathbb{R}^{n_z}$ is called semi-explicit.

Here, the DAE consists of two coupled subsystems: One subsystem of the differential equations (4.4) and one subsystem of the algebraic constraints (4.5). Initial values for a semi-explicit DAE $w(t_0) = w_0,\ z(t_0) = z_0$ have to be *consistent*, i.e. they have to fulfill the algebraic constraints. For a semi-explicit DAE of index-1, this reads

$$0 = g(w_0, z_0).$$

The index-1 condition for semi-explicit DAEs can be expressed by the properties of the algebraic constraints.

Lemma 13. *Let be given a semi-explicit DAE (4.4-4.5) and g is differentiable with respect to w and z. If*

$$\det\left(\frac{\partial g}{\partial z}\right) \neq 0 \tag{4.6}$$

*holds in a neighbourhood of the exact solution, then the DAE is of index-*1.

Proof. Differentiating g with respect to t gives

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} g(w(t), z(t)) = \frac{\partial g}{\partial w} f(w(t), z(t)) + \frac{\partial g}{\partial z} \dot{z}.$$

Solving for $\dot{z}$ and the proof is complete. □

For the computation of a numerical approximation of a semi-explicit DAE, the existence of a uniqe solution is mandatory.

Lemma 14. *Let be given an index-*1 *semi-explicit DAE (4.4-4.5) on $I = [t_0, t_{end}]$ with consistent initial values $w_0,\ z_0$. If g is differentiable w.r.t. w and z, and $f, \frac{\partial g}{\partial w}, \frac{\partial g}{\partial z}$ are Lipschitz continuous on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_z}$ for all $t \in I$, then the DAE-IVP has a unique solution.*

Proof. The index-1 condition (4.6) guarantees the regularity of the Jacobian matrix $\frac{\partial g}{\partial z}$. From the Lipschitz continuity of the partial derivatives, we deduce the boundedness of $\left\| \left(\frac{\partial g}{\partial z} \right)^{-1} \right\|$. According to the implicit function theorem, (4.5) defines a continuous differentiable function

$$G:\ \mathcal{U}_\varepsilon(w(t)) \to \mathbb{R}^{n_z}, \qquad z(t) = G(w(t)) \qquad \text{for } t \in I \tag{4.7}$$

with $\mathcal{U}_\varepsilon(w(t)) \subset \mathbb{R}^{n_w}$ a neighbourhood of the analytical solution $w(t)$. So, the DAE (4.4-4.5) is equivalent to a system of ODEs

$$\dot{w}(t) = \varphi(w(t)) := f(w(t), G(w(t))) \tag{4.8}$$

and Lipschitz continuity of f leads to the existence of a unique solution. □

After deriving the assumption for the existence of a unique solution of a semi-explicit DAE-IVP of index-1, we continue with computation of a numerical approximation. To this end, we deduce Runge-Kutta methods for DAEs.

4.1.2 One-Step Methods for semi-explicit DAEs of index-1

We study a semi-explicit DAE (4.4-4.5) on the time interval $I = [t_0, t_{\text{end}}]$ with consistent intial values $w(t_0) = w_0,\ z(t_0) = z_0$. The DAE-IVP is of index-1 (4.6) and has a unique solution on I, c.f. Lemma 14. In this section, we briefly sketch the derivation of Runge-Kutta methods for semi-explicit DAE-IVPs of index-1, so that we can extend the multirate-methods of Section 2.2.1 to DAEs. More information about time integration of DAEs can be found in [HW02, BCP95, KM06].

We derive Runge-Kutta methods for the semi-explicit DAE-IVP (4.4-4.5) by the so-called *direct approach* or *ε-embedding*. To this end, we consider the corresponding *singular perturbed* system of ODEs

$$\begin{aligned} \dot{w}^\varepsilon &= f(w^\varepsilon, z^\varepsilon) \\ \varepsilon \dot{z}^\varepsilon &= g(w^\varepsilon, z^\varepsilon) \end{aligned}$$

with $\varepsilon > 0$ and initial values $w_0^\varepsilon = w_0,\ z_0^\varepsilon = z_0$. The analytical solution is denoted by $w^\varepsilon(t), z^\varepsilon(t)$. We apply an implicit s-stage Runge-Kutta method with regular coefficient matrix $A = (a_{ij})_{i,j=1}^s$, stage vector $(c_i)_{i=1}^s$ and weight vector $(b_j)_{j=1}^s$. The transition $\varepsilon \to 0$ will lead to a Runge-method for the semi-explicit DAE.

The time step $t_n \to t_{n+1} = t_n + h$ for the singular perturbed ODE-IVP reads

$$w_{n+1}^\varepsilon = w_n^\varepsilon + h \sum_{j=1}^s b_j f(W_{nj}^\varepsilon, Z_{nj}^\varepsilon) \tag{4.9}$$

$$z_{n+1}^\varepsilon = z_n^\varepsilon + h \sum_{j=1}^s b_j \frac{1}{\varepsilon} g(W_{nj}^\varepsilon, Z_{nj}^\varepsilon) \tag{4.10}$$

with approximations $w_n^\varepsilon \approx w^\varepsilon(t_n), w_{n+1}^\varepsilon \approx w^\varepsilon(t_{n+1})$, analogously for z^ε. The intermediate stage

values $W_{ni}^{\varepsilon} \approx w^{\varepsilon}(t_n + c_i h)$, $Z_{ni}^{\varepsilon} \approx z^{\varepsilon}(t_n + c_i h)$ are given by

$$\frac{W_{ni}^{\varepsilon} - w_n^{\varepsilon}}{h} = \sum_{j=1}^{s} a_{ij} f(W_{nj}^{\varepsilon}, Z_{nj}^{\varepsilon}), \qquad \frac{Z_{ni}^{\varepsilon} - z_n^{\varepsilon}}{h} = \sum_{j=1}^{s} a_{ij} \frac{1}{\varepsilon} g(W_{nj}^{\varepsilon}, Z_{nj}^{\varepsilon})$$

for $i = 1, \ldots, s$. Multiplying the second equation with $A^{-1} = (\alpha_{ki})_{k,i=1}^{s}$ leads to

$$\sum_{i=1}^{s} \alpha_{ki} \frac{Z_{ni}^{\varepsilon} - z_n^{\varepsilon}}{h} = \frac{1}{\varepsilon} g(W_{nk}^{\varepsilon}, Z_{nk}^{\varepsilon}) \tag{4.11}$$

$$\Leftrightarrow \quad \varepsilon \sum_{i=1}^{s} \alpha_{ki} \frac{Z_{ni}^{\varepsilon} - z_n^{\varepsilon}}{h} = g(W_{nk}^{\varepsilon}, Z_{nk}^{\varepsilon}) = 0 \quad \text{for} \quad \varepsilon \to 0 \tag{4.12}$$

for $k = 1, \ldots, s$. We insert (4.11) into (4.10) and derive

Definition 10 (Runge-Kutta Methods for Semi-Explicit DAEs of Index-1)**.** *Let be given an implicit s-stage Runge-Kutta method with parameters A, b, c and regular coefficient matrix $A^{-1} = (\alpha_{ki})_{k,i=1}^{s}$. This method is applied to the semi-explicit DAE (4.4-4.5) on $[t_0, t_{end}]$ with consistent initial values w_0, z_0, fulfilling the index-*1 *condition* (4.6) *and the assumptions of Lemma 14. Then, the approximations w_{n+1}, z_{n+1} at $t_{n+1} = t_n + h$ are given by*

$$W_{ni} = w_n + h \sum_{j=1}^{s} a_{ij} f(W_{nj}, Z_{nj}), \qquad i = 1, \ldots, s \tag{4.13}$$

$$0 = g(W_{ni}, Z_{ni}), \qquad i = 1, \ldots, s \tag{4.14}$$

$$w_{n+1} = w_n + h \sum_{j=1}^{s} b_j f(W_{nj}, Z_{nj}) \tag{4.15}$$

$$z_{n+1} = \left(1 - \sum_{j=1}^{s} \sum_{i=1}^{s} b_j \alpha_{ji}\right) z_n + \sum_{j=1}^{s} \sum_{i=1}^{s} b_j \alpha_{ji} Z_{ni} \tag{4.16}$$

with intermediate stage values $W_{ni} \approx w(t_n + c_i h)$, $Z_{ni} \approx z(t_n + c_i h)$.

We point out, that generally the algebraic constraints at t_{n+1} are not fulfilled, i.e. $g(w_{n+1}, z_{n+1}) \neq 0$. To overcome with, one can replace (4.16) by the condition $g(w_{n+1}, z_{n+1}) = 0$ or by using a

Definition 11 (Stiffly Accurate Runge-Kutta Method)**.** *An implicit, s-stage Runge-Kutta method with parameters A, b, c is called stiffly accurate if*

$$a_{sj} = b_j, \tag{4.17}$$

hold for $j = 1, \ldots, s$

As direct consequence, the approximations computed by a stiffly accurate, s-stage Runge-Kutta method, are given by

$$w_{n+1} = W_{ns} \qquad \text{and} \qquad z_{n+1} = Z_{ns}.$$

Then, due to (4.14), the algebraic constraint at t_{n+1} is always fulfilled.

An important property of stiffly accurate Runge-Kutta methods is shown in [DHZ87,GM86]: The order of accuracy in the ODE-case is preserved also for the application to semi-explicit DAEs of index-1 (for w as well as for z).

We see, that stiffly accurate Runge-Kutta methods fit well for time integration of semi-explicit DAE-IVPs of index-1. The implicit Euler method is a simple example for a stiffly accurate Runge-Kutta scheme. The next chapter derives and investigates a multirate scheme for semi-explicit DAEs of index-1 based on the implicit Euler method.

4.2 The Multirate Implicit Euler Method for Semi-Explicit DAEs of Index-1

To derive a multirate Runge-Kutta method for DAEs, we generalize the concept of multiscale ODE-IVPs and consider a coupled system of DAEs with different dynamical behaviour (Section 4.2.1). Based on the implicit Euler method, we derive in Section 4.2.2 a multirate integration method for semi-explicit DAEs of index-1 and three different coupling approaches. We prove analytically, that the resulting scheme is consistent and convergent of order 1 (Sections 4.2.3 and 4.2.4). Numerical Simulations in Section 4.2.5 confirm the theoretical results.

This section is an extension of the results of [HBGS19].

4.2.1 Multiscale Differential-Algebraic Equations

Similarly to a multiscale partitioned ODE-IVP (2.3), we study two coupled DAEs in semi-explicit form

$$\dot{w}_F = f_F(w_F, z_F, w_S, z_S) \tag{4.18}$$
$$0 = g_F(w_F, z_F, w_S, z_S) \tag{4.19}$$
$$\dot{w}_S = f_S(w_F, z_F, w_S, z_S) \tag{4.20}$$
$$0 = g_S(w_F, z_F, w_S, z_S) \tag{4.21}$$

with $x_v : [t_0, t_{\text{end}}] \to \mathbb{R}^{n_v^x}$ while $x \in \{w, z\},\ v \in \{F, S\}$ and $t_0 < t_{\text{end}} \in \mathbb{R}$. The system provides the particular multirate behaviour: w_F, z_F are changing much faster than the slow components w_S, z_S. A set of consistent initial values $w_F(t_0) = w_{F0},\ z_F(t_0) = z_{F0},\ w_S(t_0) = w_{S0},\ z_S(t_0) = z_{S0}$ is given, such that

$$g_F(w_{F0}, z_{F0}, w_{S0}, z_{S0}) = 0 \qquad \text{and} \qquad g_S(w_{F0}, z_{F0}, w_{S0}, z_{S0}) = 0.$$

The right-hand sides of the system $f_F,\ g_F,\ f_S,\ g_S$ are assumed to be sufficiently smooth and that each subsystem itself has differential index-1

$$\det\left(\frac{\partial g_F}{\partial z_F}\right) \neq 0 \quad \text{and} \quad \det\left(\frac{\partial g_S}{\partial z_S}\right) \neq 0 \quad \forall (w_F, z_F, w_S, z_S) \in \mathcal{U}_\varepsilon(w_F(t), z_F(t), w_S(t), z_S(t)) \tag{4.22}$$

while $\mathcal{U}_\varepsilon(w_F(t), z_F(t), w_S(t), z_S(t))$ denotes an open environment around the analytical solution of the coupled DAE-system (4.18-4.21). In the first inequality $w_S(t), z_S(t)$ are seen as time-dependent input-functions to the fast subsystem and vice versa in the second inequality $w_F(t), z_F(t)$ are input-functions to the slow subsystem. Moreover, we claim the coupled DAE-system to be of

index-1

$$\det\left(\frac{\partial g}{\partial z}\right) \neq 0 \qquad \forall (w_F, z_F, w_S, z_S) \in \mathcal{U}_\varepsilon(w_F(t), z_F(t), w_S(t), z_S(t)) \tag{4.23}$$

with

$$z = \begin{pmatrix} z_F \\ z_S \end{pmatrix} \quad \text{and} \quad g(w_F, z_F, w_S, z_S) = \begin{pmatrix} g_F(w_F, z_F, w_S, z_S) \\ g_S(w_F, z_F, w_S, z_S) \end{pmatrix}.$$

The index-1 condition allows us to apply the implicit function theorem to solve the algebraic constraints (locally) for the algebraic variables

$$z_S = G_S(w_F, w_F) \quad \text{and} \quad z_F = G_F(w_F, w_S) \tag{4.24}$$

for $(w_F, z_F, w_S, z_S) \in \mathcal{U}_\varepsilon(w_F(t), z_F(t), w_S(t), z_S(t))$ and implicit functions $G_v : \mathcal{U}_\varepsilon(w_F(t), w_S(t)) \to \mathbb{R}^{n_v^z}$, $v \in F, S$. These properties allows us to integrate the system (4.18-4.21) with an implicit and stiffly-accurate Runge-Kutta method as described in Section 4.1.2. We continue with the derivation of an efficient multirate time integration method for coupled DAEs.

4.2.2 The mrIRK-1 Scheme for Semi-Explicit DAEs of Index-1

We consider the coupled DAE-system (4.18-4.21). We assume index-1 for the subsystems and the coupled system, respectively (4.22, 4.23). According to the results of Section 4.1.2 the coupled system and the subsystems can be integrated with a stiffly-accurate, implicit Runge-Kutta method. To exploit the particular dynamical behaviour of the DAE-system (4.18-4.21), we extend a multirate Runge-Kutta method for ODE-IVPs of Section 2.2.1 to the coupled system of semi-explicit DAEs of index-1. To this end, we consider the implicit Euler method with Butcher-tableau

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$$

which has a regular coefficient matrix $a_{11} = 1$ and is stiffly-accurate $a_{11} = b_1$. Therefore, it can be used for time integration of semi-explicit DAEs of index-1. The slow subsystem (4.20-4.21) is integrated with a large macro-step size H, the fast subsystem (4.18-4.19) with a smaller micro-step size $h = H/m$ for a fixed multirate factor $m \in \mathbb{N}$.

The macro-step $t_n \to t_{n+1} = t_n + H$ for the computation of the slow variables and the compound step in case of the Coupled-First-Step approach, repectively, reads

$$w^*_{Fn+k/m} = w_{Fn} + h \cdot k \cdot f_F(w^*_{Fn+k/m}, z^*_{Fn+k/m}, w_{Sn+1}, z_{Sn+1}) \tag{4.25}$$

$$0 = k \cdot g_F(w^*_{Fn+k/m}, z^*_{Fn+k/m}, w_{Sn+1}, z_{Sn+1}) \tag{4.26}$$

$$w_{Sn+1} = w_{Sn} + H \cdot f_S(w^*_{Fn+k/m}, z^*_{Fn+k/m}, w_{Sn+1}, z_{Sn+1}) \tag{4.27}$$

$$0 = g_S(w^*_{Fn+k/m}, z^*_{Fn+k/m}, w_{Sn+1}, z_{Sn+1}) \tag{4.28}$$

with coupling variables $w^*_{Fn+k/m}, z^*_{Fn+k/m}$ and $k \in \{0, 1, m\}$. k defines the coupling strategy according to Section 2.2.2:

$k = 0$ Decoupled-Slowest-First: For the computation of w_{Sn+1}, z_{Sn+1} the values of the fast subsystem are achieved by constant extrapolation, i.e. $w^*_{Fn+1} = w_{Fn}$, $z^*_{Fn+1} = z_{Fn}$.

$k = 1$ Coupled-First-Step: The first micro-step of the fast subsystem is computed coupled together with the macro-step of the slow subsystem, i.e. $w^*_{Fn+k/m} = w_{Fn+1/m},\ z^*_{Fn+k/m} = z_{Fn+1/m}$.

$k = m$ Coupled-Slowest-First: The fast and the slow subsystems are solved on the macro-step level coupled together, the approximations w^*_{Fn+1}, z^*_{Fn+1} are incaccurate and therefore refused.

We point out that for $k = 0$ the equations (4.25-4.26) become trivial and are not considered for the computation of w_{Sn+1} and z_{Sn+1}.

For the integration of the fast subsystem (4.18-4.19) the macro-step is split into micro-steps of size $h = H/m$. One micro-step $t_{n+l/m} \to t_{n+(l+1)/m} = t_{n+l/m} + h$ is given by

$$w_{Fn+(l+1)/m} = w_{Fn+l/m} + h \cdot f_F(w_{Fn+(l+1)/m}, z_{Fn+(l+1)/m}, \bar{w}_{Sn+(l+1/m)}, \bar{z}_{Sn+(l+1)/m}) \tag{4.29}$$

$$0 = g_F(w_{Fn+(l+1)/m}, z_{Fn+(l+1)/m}, \bar{w}_{Sn+(l+1/m)}, \bar{z}_{Sn+(l+1)/m}). \tag{4.30}$$

We have $l = 0, \ldots, m-1$ for $k \in \{0, m\}$ and $l = 1, \ldots, m-1$ for $k = 1$. The values of the slow subsystem $\bar{w}_{Sn+(l+1)/m} \approx w_S(t_{n+(l+1)/m})$ and $\bar{z}_{Sn+(l+1)/m} \approx z_S(t_{n+(l+1)/m})$ are achieved by linear interpolation, see equations (2.18)). For a system of DAEs, a Hermite-Interpolation (2.19) is not feasible due to the lack of the time-derivative of the algebraic variable z_S.

Another approach for $\bar{z}_{Sn+(l+1)/m}$ is the implicit definition via the non-linear equation

$$\begin{aligned} w_{Fn+(l+1)/m} &= w_{Fn+l/m} + h \cdot f_F(w_{Fn+(l+1)/m}, z_{Fn+(l+1)/m}, \bar{w}_{Sn+(l+1/m)}, \bar{z}_{Sn+(l+1)/m}) \\ 0 &= g_F(w_{Fn+(l+1)/m}, z_{Fn+(l+1)/m}, \bar{w}_{Sn+(l+1/m)}, \bar{z}_{Sn+(l+1)/m}) \\ 0 &= g_S(w_{Fn+(l+1)/m}, z_{Fn+(l+1)/m}, \bar{w}_{Sn+(l+1/m)}, \bar{z}_{Sn+(l+1)/m}). \end{aligned} \tag{4.31}$$

It is obvious that the computational effort increases by adding n^z_S equations to the nonlinear system. Nevertheless, for a small number n^z_S this formulation avoids the interpolation of algebraic variables which probably leads to a drift-off effect during the the integration of the fast subsystem. We show that both realisations of the algebraic-to-fast coupling lead to the same consistency order of the integration method.

From now on, we refer to the resulting multirate integration method (4.25-4.28) and (4.29-4.30) or (4.31) as **mrIRK1-DAE** scheme: **m**ulti**r**ate **I**mplicit **R**unge-**K**utta method of order **1** for semi-explicit **DAE**s of index-1.

4.2.3 Consistency Analysis for mrIRK1-DAE

We estimate the error that is made during one macro-step $t_n \to t_{n+1} = t_n + H$ caused by the mrIRK1-DAE method (based on m micro-steps; i.e., $H = m \cdot h$). We discuss the three introduced coupling strategies. Before deriving an expression for the error for all coupling approaches we introduce the notation and list up all assumptions.

Let $x : [t_0, t_{\text{end}}] \to \mathbb{R}^k$ denote some set of variables of the above DAE-system (4.18-4.21) and let exact initial values $x(t_n)$ be given for the macro-step $[t_n, t_{n+1}]$. At the end of the macro-step ($t = t_{n+1}$), we have a numerical approximation x_{n+1} of an analytic solution $x(t_{n+1})$ and the error notation:

$$\Delta x_{n+1} := x_{n+1} - x(t_{n+1}). \tag{4.32}$$

Hence, we assume at $t = t_n$: (for any vector norm $||\cdot||$)

$$\|\Delta w_{Fn}\| = \|\Delta z_{Fn}\| = \|\Delta w_{Sn}\| = \|\Delta z_{Sn}\| = 0. \tag{4.33}$$

For simplicity of notation, we introduce the following sloppy short-hand on the nth macro-step:

$$\|x(t)\|_\infty := \max_{\tau\in[t_n,t_{n+1}]} \|x(\tau)\|.$$

The following assumption is valid for the whole subsection.

Assumption 15. *For some $\varepsilon > 0$ and the analytic solution $(w_F(.), w_S(.), z_S(.))$ of the DAE (4.18-4.21), we define the neighbourhood at time τ*

$$\mathscr{E}(\tau) := \{(w_F, z_F, w_S, z_S) \in \mathbb{R}^{n_F^w+n_F^z+n_S^w+n_S^z} \Big| \\ \|w_F - w_F(\tau)\|, \|z_F - z_F(\tau)\|, \|w_S - w_S(\tau)\|, \|z_S - z_S(\tau)\| \le \varepsilon\}$$

and assume the following:

(i) The right-hand sides of DAE-system (4.18-4.21) f_F, g_F, f_S, g_S are sufficiently smooth and all first partial derivatives are (locally) uniformly bounded and the same holds for the second derivatives w.r.t. t. The Lipschitz constant of f_F with respect to w_S reads

$$L_{w_S}^{f_F} := \max_{\tau\in[t_n,t_{n+1}],\mathscr{E}(\tau)} \left\| \frac{\partial f_F}{\partial w_S}(w_F, z_F, w_S, z_S) \right\|, \tag{4.34}$$

and $L_{w_F}^{f_F}, L_{z_F}^{f_F}, L_{z_S}^{f_F}, L_{w_F}^{f_S}, L_{z_F}^{f_S}, L_{w_S}^{f_S}, L_{z_S}^{f_S}$ are defined analogously.

(ii) For the DAE-system (4.18-4.21), the implicit functions G_F and G_S (4.24) shall exist and be unique on $[t_n, t_{n+1}]$. G_F and G_S shall be sufficiently smooth and the partial derivatives shall be uniformly bounded. The corresponding Lipschitz constant reads:

$$L_{w_S}^{G_F} := \max_{\tau\in[t_n,t_{n+1}],\mathscr{E}(\tau)} \left\| \frac{\partial G_F}{\partial w_S}(w_F, w_S) \right\| \tag{4.35}$$

and $L_{w_F}^{G_F}, L_{w_F}^{G_S}, L_{w_S}^{G_S}$ analogously.

Preparing the accuracy analysis of the mrIRK-1 scheme for DAEs, we state and proof the following formulation of the mean value theorem:

Lemma 16. *Let be given a function*

$$F : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_3} \times \mathbb{R}^{n_4} \to \mathbb{R}^{n_5}$$

which is differentiable with respect to all input arguments and its Lipschitz constants are given by L_1, L_2, L_3, L_4. For $x_1, \tilde{x}_1 \in \mathbb{R}^{n_1}$, $x_2, \tilde{x}_2 \in \mathbb{R}^{n_2}$, $x_3, \tilde{x}_3 \in \mathbb{R}^{n_3}$, $x_4, \tilde{x}_4 \in \mathbb{R}^{n_4}$ we can estimate

$$\|F(x_1, x_2, x_3, x_4) - F(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4)\| \\ \le L_1\|x_1 - \tilde{x}_1\| + L_2\|x_2 - \tilde{x}_2\| + L_3\|x_3 - \tilde{x}_3\| + L_4\|x_4 - \tilde{x}_4\|. \tag{4.36}$$

Proof. By the mean value theorem, we can write

$$F(x_1,x_2,x_3,x_4)-F(\tilde{x}_1,\tilde{x}_2,\tilde{x}_3,\tilde{x}_4)=\int_0^1 \frac{\partial F}{\partial x_1}(\Theta(\sigma))\mathrm{d}\sigma(x_1-\tilde{x}_1)+\int_0^1 \frac{\partial F}{\partial x_2}(\Theta(\sigma))\mathrm{d}\sigma(x_2-\tilde{x}_2)$$
$$+\int_0^1 \frac{\partial F}{\partial x_3}(\Theta(\sigma))\mathrm{d}\sigma(x_3-\tilde{x}_3)+\int_0^1 \frac{\partial F}{\partial x_4}(\Theta(\sigma))\mathrm{d}\sigma(x_4-\tilde{x}_4) \quad (4.37)$$

with evalutation at

$$\Theta(\sigma):=\begin{pmatrix}\tilde{x}_1+\sigma(x_1-\tilde{x}_1)\\ \tilde{x}_2+\sigma(x_2-\tilde{x}_2)\\ \tilde{x}_3+\sigma(x_3-\tilde{x}_3)\\ \tilde{x}_4+\sigma(x_4-\tilde{x}_4)\end{pmatrix}.$$

We take norms on both sides of (4.37) and set

$$L_i:=\max_{\mathcal{U}_\varepsilon}\left\|\frac{\partial F}{\partial x_i}(\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{x}_4)\right\|,\quad i=1,\dots,4$$

while the maximum is taken with respect to

$$\mathcal{U}_\varepsilon=\{\|x_1-\bar{x_1}\|,\|x_2-\bar{x_2}\|,\|x_3-\bar{x_3}\|,\|x_4-\bar{x_4}\|\le\varepsilon\}$$

and we end up with the statement of the Lemma. □

The statement of this Lemma will be used in several proofs of the consistency analysis of the mrIRK1-DAE scheme.

Except for the first step in the Coupled-First-Step strategy, the computation of the fast components is the same. Thus, we start the error estimation for the fast subsystem.

Accuracy of the Fast Components

For $k\in\{0,m\}$ the estimation of the errors of the fast subsystems coincides. For $k=1$ the analysis of the error is done in a similar way but based on a different expression for the first micro-step. All three coupling strategies lead to the same order of consistency and therefore we do not present the details for $k=1$.

The following Lemma provides an estimate for the error of the fast subsystem in one macro-step $[t_n,t_{n+1}]$.

Lemma 17. *Let be given an index-*1 *DAE-IVP on* $[t_n,t_{n+1}]$ *(4.18-4.21), which fulfils Ass. 15. Let the approximation* w_{Fn+1}, z_{Fn+1}, w_{Sn+1}, z_{Sn+1} *be computed by the mrIRK1-DAE scheme (4.25-4.28) and (4.29-4.30) with macro-step size* H *and micro-step size* $h=H/m$ $(m\in\mathbb{N})$*. If the micro-step size is restricted to*

$$0<1-h(L_{w_F}^{f_F}+L_{z_F}^{f_F}L_{w_F}^{G_F})<1, \quad (4.38)$$

then the error in the fast subsystem after one macro-step $t_n \to t_n + H$ can be bounded by

$$\|\Delta w_{Fn+1}\| \le D_1 \cdot \Big\{ \frac{H^2}{2} \max_{\tau\in[t_n,t_{n+1}]} \|\ddot{w}_F(\tau)\| + h \sum_{l=1}^{m} \big((L_{w_S}^{f_F} + L_{z_F}^{f_F} L_{w_S}^{G_F}) \|\Delta \bar{w}_{Sn+l/m}\| + L_{z_S}^{f_F} \|\Delta \bar{z}_{Sn+l/m}\| \big) \Big\}$$

with a constant $D_1 \ge \Big(1 - h(L_{w_F}^{f_F} + L_{z_F}^{f_F} L_{w_F}^{G_F})\Big)^{-m} > 0$ *and the coupling errors* $\Delta \bar{w}_{Sn+l/m}$, $\Delta \bar{z}_{Sn+l/m}$.

Proof. We start with an estimation of $\Delta z_{Fn+(l+1)/m}$ in one micro-step $t_{n+l/m} \to t_{n+(l+1)/m}$:

$$\begin{aligned} \|\Delta z_{Fn+(l+1)/m}\| &= \|G_F(w_{Fn+(l+1)/m}, \bar{w}_{Sn+(l+1)/m}) - G_F(w_F(t_{n+(l+1)/m}), w_S(t_{n+(l+1)/m}))\| \\ &\le L_{w_F}^{G_F} \|\Delta w_{Fn+(l+1)/m}\| + L_{w_S}^{G_F} \|\Delta \bar{w}_{Sn+(l+1)/m}\| \end{aligned} \tag{4.39}$$

For the fast changing, differential variable we have

$$\begin{aligned} \Delta w_{Fn+(l+1)/m} = \underbrace{w_{Fn+l/m} - w_F(t_{n+l/m})}_{=\Delta w_{Fn+l/m}} &+ \big(-w_F(t_{n+(l+1)/m}) + w_F(t_{n+l/m})\big) \\ &+ h f_F(w_F(t_{n+(l+1)/m}), z_F(t_{n+(l+1)/m}), w_S(t_{n+(l+1)/m}), z_S(t_{n+(l+1)/m})) \\ &+ h f_F(w_{Fn+(l+1)/m}, z_{Fn+(l+1)/m}, \bar{w}_{Sn+(l+1)/m}, \bar{z}_{Sn+(l+1)/m}) \\ &- h f_F(w_F(t_{n+(l+1)/m}), z_F(t_{n+(l+1)/m}), w_S(t_{n+(l+1)/m}), z_S(t_{n+(l+1)/m})). \end{aligned}$$

The local truncation error of the single-rate implicit Euler method is defined as

$$\begin{aligned} \delta_{n+l/m} = w_F(t_{n+l/m}) &+ h f_F(w_F(t_{n+(l+1)/m}), z_F(t_{n+(l+1)/m}), w_S(t_{n+(l+1)/m}), z_S(t_{n+(l+1)/m})) \\ &- w_F(t_{n+(l+1)/m}) \end{aligned}$$

and can be estimated by

$$\|\delta_{n+l/m}\| \le \frac{h^2}{2} \max_{\tau\in[t_{n+l/m}, t_{n+(l+1)/m}]} \|\ddot{w}_F(\tau)\|.$$

Applying Lemma 16, we get

$$\begin{aligned} \|\Delta w_{Fn+(l+1)/m}\| \le \|\Delta w_{Fn+l/m}\| + \tfrac{h^2}{2}\|\ddot{w}_F(t)\|_\infty + h\Big(L_{w_F}^{f_F} \|\Delta w_{Fn+(l+1)/m}\| + L_{z_F}^{f_F} \|\Delta z_{Fn+(l+1)/m}\| \\ + L_{w_S}^{f_F} \|\Delta \bar{w}_{Sn+(l+1)/m}\| + L_{z_S}^{f_F} \|\Delta \bar{z}_{Sn+(l+1)/m}\| \Big). \end{aligned}$$

For $\|\Delta z_{Fn+(l+1)/m}\|$ we insert the result of (4.39). Summing all micro-steps ($l = 0, 1, \ldots, m-1$), using exact IVs at $t = t_n$ (4.33), we arrive at the statement of the lemma. □

It remains to estimate $\Delta \bar{w}_{Sn+l/m}$, $\Delta \bar{z}_{Sn+l/m}$ for all $l = 0, 1, \ldots, m-1$. The following lemma gives a corresponding bound:

Lemma 18. *Under the same settings and assumptions as in Lemma 17, the coupling errors can be bounded by*

a) $\|\Delta \bar{w}_{Sn+l/m}\| \le \frac{1}{2} l h^2 (m-l) \|\ddot{w}_S(\tau)\| + \frac{l}{m} \|\Delta w_{Sn+1}\|$ *for some* $\tau \in [t_n, t_{n+1}]$,

b) $\|\Delta\bar{z}_{Sn+l/m}\| \le \frac{1}{2}lh^2(m-l)\|\ddot{z}_S(\tau)\| + \frac{l}{m}\|\Delta z_{Sn+1}\|$ *for some* $\tau \in [t_n, t_{n+1}]$
if $\bar{z}_{Sn+l/m}$ *is achieved by linear interpolation* (2.18)*,*

c) $\|\Delta\bar{z}_{Sn+l/m}\| \le L^{G_S}_{w_F}\|\Delta w_{Fn+l/m}\| + L^{G_S}_{w_S}\|\Delta\bar{w}_{Sn+l/m}\|$
if the formulation based on the algebraic constraint (4.31) *is used.*

Proof. a) It holds:

$$\begin{aligned}\Delta\bar{w}_{Sn+l/m} &= w_S(t_{n+l/m}) - \left(\tfrac{m-l}{m}w_{Sn} + \tfrac{l}{m}w_{Sn+1}\right)\\ &= w_S(t_{n+l/m}) - \left(\tfrac{m-l}{m}w_{Sn} + \tfrac{l}{m}w_S(t_{n+1})\right) - \tfrac{l}{m}\Delta w_{Sn+1}.\end{aligned}$$

Then, an error estimation for linear interpolation yields a).

b) Analogous to a).

c) We have

$$\|\Delta\bar{z}_{Sn+l/m}\| = \|G_S(w_{Fn+l/m}, \bar{w}_{Sn+l/m}) - G_S(w_F(t_{n+l/m}), w_S(t_{n+l/m}))\|.$$

Lemma 16 yields the statement. □

To estimate Δw_{Fn+1} in terms of Δw_{Sn+1} and Δz_{Sn+1}, we combine the previous lemmas and have as direct consequence:

Proposition 19. *Under the same settings and assumptions as in Lemma 17, the error* Δw_{Fn+1} *can be bounded (using linear interpolation for* $\bar{w}_S$*):*

i) for $\bar{z}_S$ *obtained by linear interpolation* (2.18)

$$\begin{aligned}\|\Delta w_{Fn+1}\| \le D_1 \cdot \Big[&\tfrac{H^2}{2m}\|\ddot{w}_F(t)\|_\infty + \tfrac{H+h}{2}(L^{f_F}_{w_S} + L^{f_F}_{z_F}L^{G_F}_{w_S})\Big(\tfrac{H^2+Hh}{6}\|\ddot{w}_S(\tau)\|_\infty + \|\Delta w_{Sn+1}\|\Big)\\ &+ \tfrac{H+h}{2}L^{f_F}_{z_S}\Big(\tfrac{H^2+Hh}{6}\|\ddot{z}_S(t)\|_\infty + \|\Delta z_{Sn+1}\|\Big)\Big];\end{aligned}$$

ii) for $\bar{z}_S$ *computed by the non-linear equation* (4.31) *and h restricted to*

$$0 < 1 - h(L^{f_F}_{w_F} + L^{f_F}_{z_F}L^{G_F}_{w_F} + L^{f_F}_{z_S}L^{G_S}_{w_F}) < 1, \tag{4.40}$$

then we have the bound

$$\begin{aligned}\|\Delta w_{Fn+1}\| \le D_2 \cdot \Big[&\tfrac{H^2}{2m}\|\ddot{w}_F(t)\|_\infty\\ &+ \tfrac{H+h}{2}(L^{f_F}_{w_S} + L^{f_F}_{z_F}L^{G_F}_{w_S} + L^{f_F}_{z_S}L^{G_S}_{w_S})\Big(\tfrac{H^2}{6}\|\ddot{w}_S(t)\|_\infty + \|\Delta w_{Sn+1}\|\Big)\Big]\end{aligned}$$

with constant $D_2 \ge \left(1 - h(L^{f_F}_{w_F} + L^{f_F}_{z_F}L^{G_F}_{w_F} + L^{f_F}_{z_S}L^{G_S}_{w_F})\right)^{-m} > 0.$

Next, we provide estimations for $\|\Delta w_{Sn+1}\|$ and $\|\Delta z_{Sn+1}\|$ for all coupling approaches.

Accuracy of the Slow Components

The derivation of an error bound for the slow components is done in two steps: we start with an estimation for the algebraic variables, then the slow differential variables are estimated.

Lemma 20. *Let be given an index-1 coupled DAE-IVP (4.18-4.21) fulfilling Ass. 15. Let the approximation w_{Sn+1}, z_{Sn+1} be computed by the mrIRK1-DAE scheme (4.25-4.28) with macro-step size H. Let the coupling terms be denoted by $w^*_{Fn+k/m}, z^*_{Fn+k/m}$. Then the error in z_S can be bounded by*

$$\|\Delta z_{Sn+1}\| \leq \tfrac{m-k}{m} H L^{G_S}_{w_F} \max_{\tau\in[t_{n+k/m},t_{n+1}]} \|\dot{w}_F(\tau)\| + L^{G_S}_{w_F}\|\Delta w^*_{Fn+k/m}\| + L^{G_S}_{w_S}\|\Delta w_{Sn+1}\| \tag{4.41}$$

with $\tau \in [t_n, t_n+H]$ and Lipschitz constants L^{G_F}, L^{G_S}.

Proof. Solving the algebraic constraint (4.24), we can write for the local error

$$\Delta z_{Sn+1} = G_S\big(w^*_{Fn+k/m}, w_{Sn+1}\big) - G_S\big(w_F(t_{n+1}), w_S(t_{n+1})\big).$$

Applying Lemma 16, we obtain

$$\begin{aligned}\|\Delta z_{Sn+1}\| &\leq L^{G_S}_{w_F}\|w^*_{Fn+k/m} - w_F(t_{n+1})\| + L^{G_S}_{w_S}\|\Delta w_{Sn+1}\| \\ &\leq L^{G_S}_{w_F}\|w_F(t_{n+k/m}) - w_F(t_{n+1})\| + L^{G_S}_{w_F}\|\Delta w^*_{Fn+k/m}\| + L^{G_S}_{w_S}\|\Delta w_{Sn+1}\|\end{aligned}$$

The mean value theorem completes the proof. □

The estimation for the error in w_S differs between the coupled approaches $k \in \{1, m\}$ and the Decoupled-Slowest-First approach $k = 0$. First, we derive an upper bound for the latter one.

Proposition 21. *We consider the Decoupled-Slowest-First approach. Under the same settings and assumptions as in Lemma 20 (for $k = 0$) and a restricted macro-step size H, such that*

$$0 < 1 - H(L^{f_S}_{w_S} + L^{f_S}_{z_S} L^{G_S}_{w_S}) < 1 \tag{4.42}$$

holds, the error in w_S is bounded by

$$\|\Delta w_{Sn+1}\| \leq \frac{H^2}{1 - H(L^{f_S}_{w_S} + L^{f_S}_{z_S} L^{G_S}_{w_S})} \left(\left(L^{f_S}_{w_S} + L^{f_S}_{z_S} L^{G_S}_{w_S}\right) \|\dot{w}_S(\tau)\|_\infty + \tfrac{1}{2}\|\ddot{w}_S(\tau)\|_\infty \right). \tag{4.43}$$

Then, the Decoupled-Slowest-First mrIRK-1 applied to the DAE-IVP (4.18-(4.21)) is of consistency order 1 in the differential variables w_F and w_S. The error in the algebraic variables z_F, z_S is in $\mathcal{O}(H)$.

Proof. By Taylor expansion of $w_S(t_{n+1})$ with expansion point t_n, we obtain

$$\begin{aligned}\Delta w_{Sn+1} &= w_{Sn} + H \cdot f_S(w_{Fn}, z_{Fn}, w_{Sn+1}, z_{Sn+1}) - \left(w_S(t_n) + H \cdot \dot{w}_S(t_n) + \tfrac{H^2}{2}\ddot{w}_S(\tau) \right) \\ &= H \cdot \left(f_S\left(w_{Fn}, z_{Fn}, w_{Sn+1}, z_{Sn+1}\right) - f_S\left(w_F(t_n), z_F(t_n), w_S(t_n), z_S(t_n)\right)\right) - \tfrac{H^2}{2}\ddot{w}_S(\tau)\end{aligned}$$

for some $\tau \in [t_n, t_n + H]$. Applying Lemma 16 and the mean value theorem, we get

$$\begin{aligned}\|\Delta w_{Sn+1}\| &\leq H\left(L_{w_S}^{f_S}\|w_{Sn+1} - w_{Sn}\| + L_{z_S}^{f_S}\|z_{Sn+1} - z_{Sn}\|\right) + \tfrac{H^2}{2}\|\ddot{w}_S(\tau)\|_\infty \\ &\leq H\left(L_{w_S}^{f_S}\|\Delta w_{Sn+1}\| + H \cdot L_{w_S}^{f_S}\|\dot{w}_S(\tau)\|_\infty + L_{z_S}^{f_S}\|G_S(w_{Fn}, w_{Sn+1}) - G_S(w_{Fn}, w_{Sn})\|\right) \\ &\quad + \tfrac{H^2}{2}\|\ddot{w}_S(\tau)\|_\infty\end{aligned}$$

Using again Lemma 16 and the proof is complete. □

Next, it is to investigate the accuracy of the slow, differential variables w_S for $k \in \{1, m\}$. Here, the approximation w_{Sn+1} is computed together with the coupling variable $w^*_{Fn+l/m}$ and therefore also the errors Δw_{Sn+1}, $\Delta w^*_{Fn+l/m}$ depend on each other. The next lemma gives estimates for both.

Lemma 22. *We consider the Coupled-First-Step and the Coupled-Slowest-First approacht, i.e. $k \in \{1, m\}$. Under the same settings and assumptions as in Lemma 20, the error in the differential variables for the Macro-Step ($k = m$) and the Compound-Step ($k = 1$) can be bounded as follows:*

$$M(H\tfrac{k}{m}, H)\begin{pmatrix}\|\Delta w^*_{Fn+k/m}\| \\ \|\Delta w_{Sn+1}\|\end{pmatrix} \leq \begin{pmatrix}R_F \\ R_S\end{pmatrix}, \tag{4.44}$$

with

$$\begin{aligned}M(H_1, H_2) &:= \begin{pmatrix}1 - H_1(L_{w_F}^{f_F} + L_{z_F}^{f_F}L_{w_F}^{G_F} + L_{z_S}^{f_F}L_{w_F}^{G_S}) & -H_1(L_{w_S}^{f_F} + L_{z_F}^{f_F}L_{w_S}^{G_F} + L_{z_S}^{f_F}L_{w_S}^{G_S}) \\ -H_2(L_{w_F}^{f_S} + L_{z_F}^{f_S}L_{w_F}^{G_F} + L_{z_S}^{f_S}L_{w_F}^{G_S}) & 1 - H_2(L_{w_S}^{f_S} + L_{z_F}^{f_S}L_{w_S}^{G_F} + L_{z_S}^{f_S}L_{w_S}^{G_S})\end{pmatrix} \\ R_F &= H^2\left(\tfrac{k(m-k)}{m^2}\left(L_{w_S}^{f_F} + L_{z_F}^{f_F}L_{w_S}^{G_F} + L_{z_S}^{f_F}L_{w_S}^{G_S}\right)\|\dot{w}_S\|_\infty + \tfrac{2k(m-k)}{m}L_{z_S}^{f_F}L_{w_F}^{G_S}\|\dot{w}_F\|_\infty + \tfrac{k^2}{m^2}\|\ddot{w}_F\|_\infty\right) \\ R_S &= H^2\left(\tfrac{k(m-k)}{m^2}\left(L_{w_F}^{f_S} + L_{z_F}^{f_S}L_{w_F}^{G_F} + L_{z_S}^{f_S}L_{w_F}^{G_S}\right)\|\dot{w}_F\|_\infty + \tfrac{2k(m-k)}{m}L_{z_F}^{f_S}L_{w_S}^{G_F}\|\dot{w}_S\|_\infty + \tfrac{1}{2}\|\ddot{w}_S\|_\infty\right)\end{aligned}$$

The inequality in (4.44) *has to be understood componentwise.*

Proof. For Δw_{Sn+1}, we add $\pm\big[w_S(t_n) - Hf_S\big(w_F(t_{n+1}), z_F(t_{n+1}), w_S(t_{n+1}), z_S(t_{n+1})\big)\big]$. By Lemma 16, we deduce

$$\begin{aligned}\|\Delta w_{Sn+1}\| &\leq \left\|\int_0^H \tau\ddot{w}(t_n + \tau)d\tau\right\| + H\big[L_{w_F}^{f_S}\|w^*_{Fn+k/m} - w_F(t_{n+1})\| + L_{z_F}^{f_S}\|z^*_{Fn+k/m} - z_F(t_{n+1})\| \\ &\qquad + L_{w_S}^{f_S}\|\Delta w_{Sn+1}\| + L_{z_S}^{f_S}\|\Delta z_{Sn+1}\|\big] \\ &\leq \tfrac{H^2}{2}\|\ddot{w}_S\|_\infty + H\big[L_{w_F}^{f_S}\|\Delta w^*_{Fn+k/m}\| + L_{z_F}^{f_S}\|\Delta z^*_{Fn+k/m}\| + L_{w_S}^{f_S}\|\Delta w_{Sn+1}\| + L_{z_S}^{f_S}\|\Delta z_{Sn+1}\|\big] \\ &\quad + \tfrac{m-k}{m}H^2\big[L_{w_F}^{f_S}\|\dot{w}_F\|_\infty + L_{z_F}^{f_S}\|\dot{z}_F\|_\infty\big].\end{aligned}$$

The estimation for $\|\Delta z_{Sn+1}\|$ is given in Lemma 20. $\dot{z}_F$ can be solved according to (4.24). For $\Delta z^*_{Fn+k/m}$ we apply Lemma 16 and the mean value theorem and get

$$\|\Delta z^*_{Fn+k/m}\| \leq L_{w_F}^{G_F}\|\Delta w^*_{Fn+k/m}\| + L_{w_S}^{G_F}\|\Delta w_{Sn+1}\| + H\tfrac{m-k}{m}L_{w_S}^{G_F}\|\dot{w}_S\|_\infty. \tag{4.45}$$

Analogously, one can deduce the estimate for Δw^*_{Fn+1}. Combining all estimations, we arrive at the statement of the Lemma. □

To solve the estimate (4.44) for the error in the differential variables, we need that $M(H_1, H_2)$ is an M-matrix in $\mathbb{R}^{2x2}$. In fact, for $H_1, H_2 > 0$ small enough, the diagonal entries are positive

(off-diagonals are always negative). Thus, we have

Proposition 23. *Let the same settings and assumptions apply as in Lemma 22. And the step-size H and the multirate factor m be restricted such that the M-Matrix conditions for $M(H\frac{k}{m},H)$ are fulfilled [Ple77]:*

$$\begin{gathered} H(L^{f_S}_{w_S}+L^{f_S}_{z_F}L^{G_F}_{w_S}+L^{f_S}_{z_S}L^{G_S}_{w_S})<1 \quad \textit{and} \quad H\tfrac{k}{m}(L^{f_F}_{w_F}+L^{f_F}_{z_F}L^{G_F}_{w_F}+L^{f_F}_{z_S}L^{G_S}_{w_F})<1 \\ \textit{and} \quad \det\left(M(H\tfrac{k}{m},H)\right)>0 \qquad . \end{gathered} \tag{4.46}$$

Then, the Coupled-First-Step ($k=1$) and the Coupled-Slowest-First ($k=m$) mrIRK1-DAE scheme applied to the DAE-IVP (4.18-4.21) is of consistency order 1 in the differential variables w_F, w_S and also for the algebraic variables z_F, z_S for $k=m$. For $k=1$, the error in the algebraic variables z_F, z_S is in $\mathscr{O}(H)$.

Summary

We conclude for Section 4.2.2:

Theorem 24. *For all versions of the mrIRK1-DAE method applied to the DAE-IVP (4.18-4.21) the differential variables (w_F, w_S) have consistency order 1. The algebraic variables (z_F, z_S) reach order 1 only in the Coupled-Slowest-First approach. For the other coupling approaches, $\|\Delta z_S\|$, $\|\Delta z_F\|$ are always in $\mathscr{O}(H)$. Under the additional assumption*

$$\frac{\partial}{\partial w_F}G_S(w_F,w_S)=0 \qquad \textit{and} \qquad \frac{\partial}{\partial w_S}G_F(w_F,w_S)=0 \tag{4.47}$$

for $t\in[t_0,t_{end}]$ we have order 1 also in the algebraic variable $(z_F,\ z_S)$ in all coupling approaches.

In other words, the conditions (4.47) say, that the coupling of the subsystem is realised by the differential variables only.

Proof. It only remains to show order 1 in z_S and z_F for *Decoupled-Slowest-First* and *Coupled-First-Step*: Since (4.47), we have $L^{G_F}_{w_S}=L^{G_S}_{w_F}=0$ in (4.41) and (4.45) and we end up with

$$\|\Delta z_S\|=\mathscr{O}(\|\Delta w_S\|) \qquad \text{and} \qquad \|\Delta z_F\|=\mathscr{O}(\|\Delta w_F\|).$$

□

Remark: The slow changing variables $(w_S,\ z_S)$ of a multirate DAE-IVP depends only weakly on w_F, therefore $\left\|\frac{\partial}{\partial w_F}G(w_F,w_S)\right\|$ is small and can be neglected in most cases.

Lemma 18 and Proposition 19 showed, that the linear interpolation of w_S and z_S during the computation of the micro-steps does not lead to a higher order of accuracy of the multirate integration method: By a similar calculation one can show that a constant extrapolation of w_S and z_S during the computation of the micro-steps, the order of accuracy on the macro-step grid is not reduced. Nevertheless, the computational effort of the linear interpolation is comparable to constant extrapolation and leads to a higher accuracy on the micro-step grid.

Next, it is shown that the reduced consistency order in the algebraic variable does not influence the convergence of the scheme.

4.2.4 Convergence of mrIRK1-DAE

Now, we investigate the error propagation over several macro-steps. For the index-1 DAE-IVP (4.18-4.21), $(w_{Fn}, z_{Fn}, w_{Sn}, z_{Sn})$ denotes the mrIRK1-DAE approximation at t_n after n macro-steps. For any components $x = x(t)$ of the unknowns, the global error reads

$$E(x,t_n) := x_n - x(t_n).$$

We show that $E(w_F,t_n)$, $E(z_F,t_n)$, $E(w_S,t_n)$, $E(z_S,t_n)$ are in $\mathscr{O}(H)$. To this end, we recall the following theorem from [DHZ87]:

given a semi-explicit DAE-IVP of index-1 (4.18-4.21), we apply a general one-step method

$$\begin{aligned} w_{k+1} &= w_k + \hat{h} \cdot \Phi(w_k, z_k, \hat{h}), \\ z_{k+1} &= \Psi(w_k, z_k, \hat{h}) \end{aligned}$$

with $w^\top = (w_F^\top, w_S^\top)$ and $z^\top = (z_F^\top, z_S^\top)$, a constant step size $\hat{h}$, a differential update function Φ and an algebraic update function Ψ. We remark that Φ and Ψ are only formally explicit. If the method has consistency order p for the differential variables w, as well as $p-1$ for algebraic variables z and if the algebraic update function satisfies the following perturbation condition

$$\left\| \frac{\partial \Psi(w,z,0)}{\partial z} \right\| \leq \alpha < 1 \tag{4.48}$$

in a neighbourhood of the solution, then the one-step method has convergence order p. For $p = 1$ this statement holds for the mrIRK1-DAE method:

Theorem 25. *We apply the mrIRK1-DAE method to the index-*1 *DAE-IVP (4.18-4.21) fulfilling Ass. 15. We may choose any coupling variant: Coupled-Slowest-First, Decoupled-Slowest-First, Coupled-First-Step. H and m are chosen such that the corresponding step size restrictions* (4.38)*,* (4.40)*,* (4.42)*,* (4.46) *are fulfilled. Then we get for the global error*

$$E(w_F,t_n) = \mathscr{O}(H), \quad E(z_F,t_n) = \mathscr{O}(H), \quad E(w_S,t_n) = \mathscr{O}(H), \quad E(z_S,t_n) = \mathscr{O}(H).$$

Proof. We check the assumptions of the theorem from [DHZ87] (mentioned above):

One-Step Method. All discussed formulations of the mrIRK1-DAE scheme define the approximations $w_{Fn+1}, z_{Fn+1}, w_{Sn+1}, z_{Sn+1}$ at t_{n+1} after one macro step as functions of the approximations $w_{Fn}, z_{Fn}, w_{Sn}, z_{Sn}$ at t_n.

Consistency. Theorem 24 showed that we have consistency order 1 for the differential variables and at least order $\mathscr{O}(H)$ for the algebraic variables (for any variant).

Perturbation Condition (4.48). For the slow changing, algebraic variable, we have

$$\begin{aligned} z_{Sn+1} &= G_S(w^*_{Fn+k/m}, w_{Sn+1}) \\ &= G_S\big(w_{Fn} + H\tfrac{k}{m} f_F(w^*_{Fn+k/m}, z^*_{Fn+k/m}, w_{Sn+1}, z_{Sn+1}), \\ &\qquad w_{Sn} + H f_S(w^*_{Fn+k/m}, z^*_{Fn+k/m}, w_{Sn+1}, z_{Sn+1})\big) \end{aligned}$$

with $k \in \{0,1,m\}$. Let Ψ_S denote the update function for the slow, algebraic variable, then we have

$$\frac{\partial \Psi_S}{\partial z_{Fn}} = H\frac{k}{m}\frac{\partial G_S}{\partial w_F}\frac{\partial f_F}{\partial z_{Fn}} + H\frac{\partial G_S}{\partial w_S}\frac{\partial f_S}{\partial z_{Fn}}$$

We point out, that the partial derivates are computed with respect to the algebraic variable of the previous time step z_n, therefore the expression above becomes zero for $k \in \{1,m\}$. For $k = 0$ the remaining term reads

$$\frac{\partial \Psi_S}{\partial z_{Fn}} = H\frac{\partial G_S}{\partial w_S}\frac{\partial f_S}{\partial z_{Fn}}$$

which vanishes for the evaluation at $H = 0$. Analogously we can derive $\frac{\partial \Psi_S}{\partial z_{Sn}} = 0$. For the fast algebraic variable we have

$$\begin{aligned} z_{Fn+1} = G_F\big(&w_{Fn} + h\sum_{l=1}^{m} f_F(w_{Fn+l/m}, z_{Fn+l/m}, \bar{w}_{Sn+l/m}, \bar{z}_{Sn+l/m}), \\ &w_{Sn} + H f_S(w^*_{Fn+k/m}, z^*_{Fn+k/m}, w_{Sn+1}, z_{Sn+1})\big) \end{aligned}$$

and the equality $\frac{\partial \Psi_F}{\partial z_{Fn}} = 0$ is derived analogously to the slow changing, algebraic variable. The coupling terms $\bar{w}_{Sn+l/m}, \bar{z}_{Sn+l/m}$ are achieved by linear interpolation and we have

$$\frac{\partial \Psi_F}{\partial z_{Sn}} = \frac{H}{m}\frac{\partial G_S}{\partial w_F}\frac{\partial f_F}{\partial z_{Sn}}\left(\frac{m-1}{m} + \frac{m-1}{m} + \cdots + \frac{1}{m}\right).$$

Evaluation at $H = 0$ and the proof is complete. □

The following numerical simulations confirm this analytical result.

4.2.5 Numerical Results

For the numerical verification, we consider two DAE-systems.

Extended Prothero-Robinson Equation

An extended Prothero-Robinson test equation for semi-explicit DAEs [BBS14] reads in our settings as follows

$$\begin{pmatrix} \dot{w} \\ 0 \end{pmatrix} = \begin{pmatrix} A - BF & B \\ C - DF & D \end{pmatrix}\begin{pmatrix} w \\ z \end{pmatrix} + \begin{pmatrix} -A\eta(t) - B\zeta(t) + \dot{\eta}(t) \\ -C\eta(t) - D\zeta(t) \end{pmatrix} \tag{4.49}$$

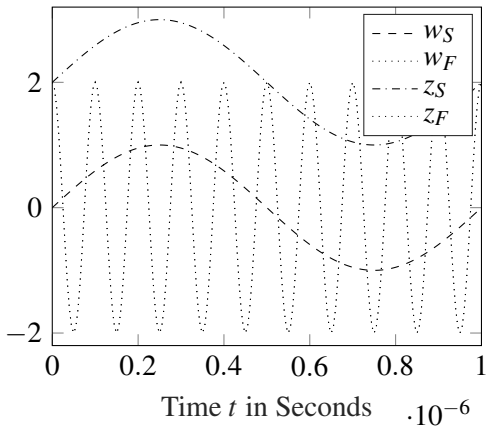


Figure 4.1: Solution of DAE-IVP (4.49).

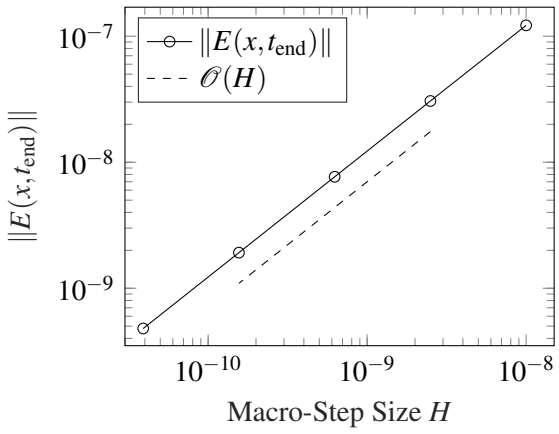


Figure 4.2: Coupled-First-Step: $\|E(x,t_{\text{end}})\|$ in $\mathscr{O}(H)$

with $w(t) = (w_S(t), w_F(t))^\top \in \mathbb{R}^2$ and $z(t) = (z_S(t), z_F(t))^\top \in \mathbb{R}^2$ and given functions η and ζ. For the simulation we choose the following data:

$$A = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix},\ B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},\ C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ D = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix},\ F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\eta(t) = (\sin(2\pi 10^6 t),\ 2\cos(2\pi 10^7 t))^\top, \qquad \zeta(t) = (2\cos(t),\ 7t)^\top. \tag{4.50}$$

Since D is regular (4.49) is of index-1 and consistent initial values are given by

$$(w_S(0), w_F(0), z_{S1}(0), z_{S2}(0))^\top = (0,2,2,2)^\top. \tag{4.51}$$

Notice that the solution of (4.49) is

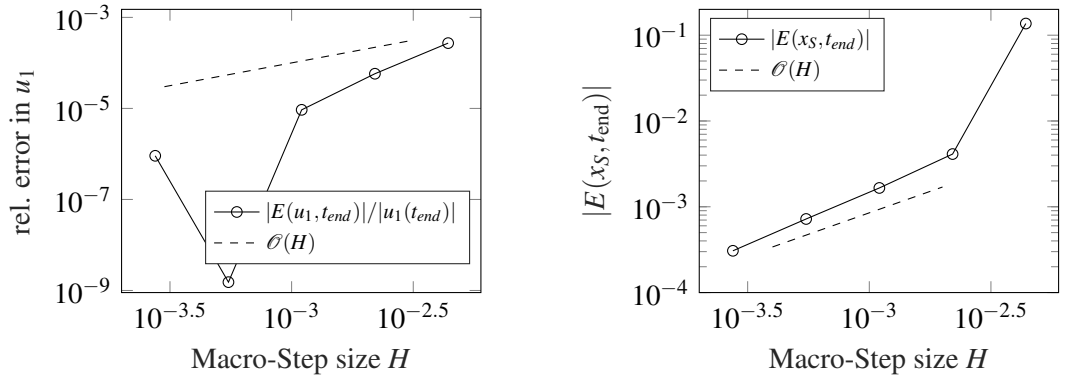
$$w(t) = \eta(t), \qquad z(t) = F\eta(t) + \zeta(t).$$

The different dynamical behaviour of the components is illustrated in Fig. 4.1.

We apply the mrIRK1-DAE method to the DAE (4.49) on $[t_0,\ t_{\text{end}}] = [0,\ 10^{-6}s]$ using all three coupling approaches. We use different macro-step sizes $H = 2^{-i} \cdot 10^{-8}$ for $i = 0,2,4,6,8$ and fixed multirate factor $m = 10$. We investigate the global error at t_{end}.

For all coupling approaches, all components of the DAE system show convergence order 1. Representative for all combinations of coupling approaches and components, we show the combined error $\|E(x,t_{\text{end}})\|$ for $x = (w_S, w_F, z_S, z_F)^\top$ for the Coupled-First-Step approach in Fig. 4.2. The Figure illustrates the convergence order of 1 for the coupled DAE-system.

A complete overview of the convergence behaviour of all components using all three coupling approaches is given in the Part A of the Appendix.



(a) Relative Error of the fast, circuit subsystem

(b) Absolute Error in the slow, field subsystem

Figure 4.3: Convergence order for the *Coupled-Slowest-First* approach for the field-circuit coupled system

Field-Circuit Coupled System

We consider the field-circuit coupled system of Section 3.4, the circuit diagram is given in Figure 3.5. We recall the system equations

$$C\dot{e}(t) = G(u_1(t) - U_{in}(t)) - \mathfrak{B}x(t) \tag{4.52}$$
$$\mathfrak{M}\dot{x}(t) = \mathfrak{F}x(t) + \mathfrak{B}u_1(t), \tag{4.53}$$

the fast-changing, ODE-subsystem (4.52) describes the node potential u_1, the slow-changing, DAE-subsystem (4.53) describes the electric field of a 2D-transformer with state space vector x. The consistency and convergence analysis for the mrIRK1-DAE scheme in section 4.2.2 can be easily adapted to linear DAEs with constant coefficients like (4.52-4.53).

We integrate system (4.52-4.53) over $[0s,\ 0.0022s]$ with the mrIRK1-DAE method using the Coupled-Slowest-First approach with different macro-step sizes
$H \in \{0.0003,\ 0.0006,\ 0.0011,\ 0.0022,\ 0.0044\}$ and multirate factor $m = 10$. The reference solution is obtained by a single-rate implicit Euler method with constant step-size $\hat{H} = 5.5 \cdot 10^{-7}$.

Figure 4.3 shows the global error of the subsystems at $t_{\text{end}} = 0.0022s$ separately. The range of the solution of the fast subsystem $u_1(t)$ is between $\pm 4.7 \cdot 10^4 V$, therefore we show in figure 4.3a the relative error. The simulation shows a slightly better behaviour than order 1. The error of the slow subsystem is also of order 1 which is illustrated in figure 4.3b.

Section Summary

We extended the multirate implicit Euler method to semi-explicit DAEs of index-1. We used three different strategies to realise the coupling between the slow and the fast subsystems. We provided assumptions on the macro-step size and the micro-step size that a consistency order 1 can be proven for all three coupling strategies and respective differential variables. For semi-explicit DAEs, the usage of the Coupled-Slowest-First approach seems favourable, since it is the only coupling strategy, where also for the algebraic variables consistency order 1 is derived. Anyway, all discussed multirate implicit Euler method have convergence order 1 for semi-explicit DAEs

of index-1 if the macro-step size is constant. Finally, numerical results for all coupling strategies confirm the theoretical investigations.

After deriving the mrIRK1-DAE method, we will generalize the approach to deduce an mrIRK2-DAE method based on the LobattoIIIC method using the the Coupled-Slowest-First approach.

4.3 A Second Order Multirate Runge-Kutta Method for DAEs

In Section 3.4, we briefly discussed a multirate time integration method for linear implicit DAEs of index-1 based on the LobattoIIIC scheme using the Coupled-Slowest-First approach. The method worked well for the field-circuit coupled problem and we observed a relevant gain of efficiency compared to the classical single-rate LobattoIIIC method. Nevertheless, a detailed consistency and convergence analysis is still missing and will be made up in the following: After adapting the multirate LobattoIIIC method for semi-explicit DAEs of index-1 (Section 4.3.1) we deduce the order of consistency 2 of the integration scheme (Section 4.3.2). Analogue to Section 4.2.4 the Convergence of the second order multirate scheme is proven (Section 4.3.3). The convergence order is illustrated by a numerical example at the end in the Section 4.3.4

The problem setting is the same as for the mrIRK1-DAE scheme in Section 4.2.1. We consider the partitioned DAE-IVP (4.18-4.21) with consistent initial values at t_0. For the entire section we assume that the coupled DAE-IVP fulfils the index-1 condition for the coupled system (4.23) and for both subsystems (4.22). Therefore, we can write the algebraic variable as a function of the differential ones in a neighbourhood of the analytical solution (4.24).

In Section 3.4 we applied the multirate LobattoIIIC-method to a linear implicit DAE with constant coefficients of index-1. Using techniques from linear algebra such a DAE can be transformed to semi-explicit form [BCP95].

4.3.1 The mrIRK2-DAE Scheme based on LobattoIIIC

For the index-1 DAE-IVP (4.18-4.21) we define a multirate time integration method based on the LobattoIIIC integration scheme. In Section 4.2.3 we showed, that only the Coupled-Slowest-First approach preserves the order of consistency for the algebraic variables. Therefore, we focus on this coupling approach in the following.

We recall the Butcher-Tableau of the LobattoIIIC method

$$\begin{array}{c|cc} 0 & \frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}.$$

Its coefficient matrix is regular and the method is stiffly accurate $a_{2,i} = b_i,\ i = 1,2$. Hence, it can be used for time integration of semi-explicit DAEs of index-1.

The macro-step $t_n \to t_{n+1} = t_n + H$ for the coupled DAE-IVP (4.18-4.21) reads

$$\begin{aligned}
W^*_{F1} &= w_{Fn} + \tfrac{H}{2}\left(f_F(W^*_{F1}, Z^*_{F1}, W_{S1}, Z_{S1}) - f_F(W^*_{F2}, Z^*_{F2}, W_{S2}, Z_{S2})\right)\\
0 &= g_F(W^*_{F1}, Z^*_{F1}, W_{S1}, Z_{S1})\\
W^*_{S1} &= w_{Sn} + \tfrac{H}{2}\left(f_S(W^*_{F1}, Z^*_{F1}, W_{S1}, Z_{S1}) - f_S(W^*_{F2}, Z^*_{F2}, W_{S2}, Z_{S2})\right)\\
0 &= g_S(W^*_{F1}, Z^*_{F1}, W_{S1}, Z_{S1})\\
W^*_{F2} &= w_{Fn} + \tfrac{H}{2}\left(f_F(W^*_{F1}, Z^*_{F1}, W_{S1}, Z_{S1}) + f_F(W^*_{F2}, Z^*_{F2}, W_{S2}, Z_{S2})\right)\\
0 &= g_F(W^*_{F2}, Z^*_{F2}, W_{S2}, Z_{S2})\\
W^*_{S2} &= w_{Sn} + \tfrac{H}{2}\left(f_S(W^*_{F1}, Z^*_{F1}, W_{S1}, Z_{S1}) + f_S(W^*_{F2}, Z^*_{F2}, W_{S2}, Z_{S2})\right)\\
0 &= g_S(W^*_{F2}, Z^*_{F2}, W_{S2}, Z_{S2})
\end{aligned} \tag{4.54}$$

with stage values $W_{S1} \approx w_S(t_n)$, $W_{S2} \approx w_S(t_{n+1})$, $W^*_{F1}, Z^*_{F1}, Z_{S1}, W^*_{F2}, Z^*_{F2}, Z_{S2}$ are defined analogously. $W^*_{F1}, Z^*_{F1}, W^*_{F2}, Z^*_{F2}$ denote the coupling terms and are later refused. The approximation at t_{n+1} of the slow subsystem (4.20-4.21) is given by

$$w_{Sn+1} = W_{S2}, \qquad z_{Sn+1} = Z_{S2}. \tag{4.55}$$

To achieve an accurate approximation for the fast subsystemm (4.18-4.19), m micro-steps of step size $h = H/m$ are carried out. One micro-step $t_n + (l-1)h \to t_n + lh$ reads

$$\begin{aligned}
W^l_{F1} &= w_{Fn+(l-1)/m} + \tfrac{h}{2}\left(f_F(W^l_{F1}, Z^l_{F1}, \bar{W}^l_{S1}, \bar{Z}^l_{S1}) - f_F(W^l_{F2}, Z^l_{F2}, \bar{W}^l_{S2}, \bar{Z}^l_{S2})\right)\\
0 &= g_F(W^l_{F1}, Z^l_{F1}, \bar{W}^l_{S1}, \bar{Z}^l_{S1})\\
W^l_{F2} &= w_{Fn+(l-1)/m} + \tfrac{h}{2}\left(f_F(W^l_{F1}, Z^l_{F1}, \bar{W}^l_{S1}, \bar{Z}^l_{S1}) + f_F(W^l_{F2}, Z^l_{F2}, \bar{W}^l_{S2}, \bar{Z}^l_{S2})\right)\\
0 &= g_F(W^l_{F2}, Z^l_{F2}, \bar{W}^l_{S2}, \bar{Z}^l_{S2})
\end{aligned} \tag{4.56}$$

with stage values $W^l_{F1} \approx w_F(t_n + (l-1)h)$, $W^l_{F2} \approx w_F(t_n + lh)$ and Z^l_{F1}, Z^l_{F2} analogously. The coupling terms $\bar{W}_S$, $\bar{Z}_S$ are linearly interpolated values of w_{Sn}, w_{Sn+1} and z_{Sn}, z_{Sn+1}, respectively. The approximation of the fast subsystem at $t_n + lh$ is given by

$$w_{Fn+l/m} = W^l_{F2}, \qquad z_{Fn+l/m} = Z^l_{F2}. \tag{4.57}$$

For the multirate time integration method that is defined by (4.54-4.55) and (4.56-4.57) for $l = 1, \ldots, m$ we introduce the abbreviation mrIRK2-DAE (**m**ulti**r**ate **R**unge-**K**utta method of order **2** for **DAE**s).

We expect an order of consistency of 2 for the mrIRK2-DAE scheme since it is based on a second order singlerate Runge-Kutta method. A detailed proof for the order of consistency is provided in the following chapter.

4.3.2 Consistency Analysis for mrIRK2-DAE

In this section we derive the order of consistency of the mrIRK2-DAE method. To this end, we estimate the error that is made during one macro-step $t_n \to t_{n+1} = t_n + H$ using exact values at t_n. The local discretisation error on the macro-step level is denoted by

$$\Delta x_{n+1} := x_{n+1} - x(t_{n+1}) \tag{4.58}$$

for x any variable of the DAE-IVP (4.18-4.21) with the analytical solution $x(t_{n+1})$ and the numerical approximation x_{n+1} at t_{n+1} computed by the mrIRK2-DAE method. For the values at t_n we assume

$$\|\Delta w_{Fn}\| = \|\Delta z_{Fn}\| = \|\Delta w_{Sn}\| = \|\Delta z_{Sn}\| = 0. \tag{4.59}$$

The derivation of the order of consistency of the mrIRK2-DAE scheme is done in the proof of the following

Proposition 26. *Let be given an index-*1 *DAE-IVP (4.18-4.21) fullfilling Assumption 15. Let the approximations $w_{Fn+1}, z_{Fn+1}, w_{Sn+1}, z_{Sn+1}$ be computed by the mrIRK2-DAE scheme (4.54-4.55) and (4.56-4.57) for $l = 1,\dots,m$. The macro-step size H and micro-step size $h = H/m$ for a fixed $m \in \mathbb{N}$ are restricted to*

$$H(L_w^f + L_z^f L_w^G) < 1 \qquad \text{and} \qquad h(L_{w_F}^{f_F} + L_{z_F}^{f_F} L_{w_F}^{G_F}) < 1. \tag{4.60}$$

with the L_w^f, L_z^f, L_w^G the Lipschitz constants of the non-partitioned DAE-system. Then, we have

$$\|\Delta w_{Fn+1}\| = \mathcal{O}(H^3), \quad \|\Delta z_{Fn+1}\| = \mathcal{O}(H^2), \quad \|\Delta w_{Sn+1}\| = \mathcal{O}(H^3), \quad \|\Delta z_{Sn+1}\| = \mathcal{O}(H^3)$$

and the order of consistency of the mrIRK2-DAE scheme is 2*.*

For the classical, single-rate LobattoIIIC-scheme applied to semi-explicit DAEs of index-1, the local truncation error of the algebraic variable is in $\mathcal{O}(H^3)$. For the multirate case, the less accurate fast changing, algebraic variables is caused by the linear interpolation of the coupling variable as we will see in the following

Proof. According to the order of the integration scheme, we start with the
Accuracy analysis of the slow components:
The mrIRK2-DAE scheme is based on the Coupled-Slowest-First approach. In the macro-step, the coupled DAE-system is integrated with one global stepsize H. For the accuracy analysis in the macro-step we re-order the DAE-system (4.18-4.21) according to differential and algebraic variables

$$\begin{aligned}
\dot{w}_F &= \tilde{f}_F(w_F, w_S, z_F, z_S)\\
\dot{w}_S &= \tilde{f}_S(w_F, w_S, z_F, z_S)\\
0 &= \tilde{g}_F(w_F, w_S, z_F, z_S)\\
0 &= \tilde{g}_S(w_F, w_S, z_F, z_S).
\end{aligned}$$

We set $w = (w_F^\top, w_S^\top)^\top$ and $z = (z_F^\top, z_S^\top)^\top$ and the above DAE-system can be written in the following, compact form

$$\dot{w} = f(w,z), \qquad 0 = g(w,z). \tag{4.61}$$

Then, the macro-step (4.54-4.55) reads

$$\begin{aligned}
W_1 &= w_n + \tfrac{H}{2}\left(f(W_1,Z_1) - f(W_2,Z_2)\right)\\
0 &= g(W_1,Z_1)\\
W_2 &= w_n + \tfrac{H}{2}\left(f(W_1,Z_1) + f(W_2,Z_2)\right)\\
0 &= g(W_2,Z_2)
\end{aligned}$$

$$w_{n+1} = W_2 \qquad z_{n+1} = Z_2 \tag{4.62}$$

with the stage value approximations W_1, Z_1 at t_n and W_2, Z_2 at t_{n+1}.

To proof the accuracy of the slow components w_S, z_S, it is sufficient to show that $\|\Delta w_{n+1}\|$ is in $\mathscr{O}(H^3)$ and $\|\Delta z_{n+1}\|$ is in $\mathscr{O}(H^2)$. Since the LobattoIIIC scheme is stiffly accurate, we actually have $\|\Delta z_{n+1}\| = \mathscr{O}(H^3)$. For the error in the algebraic variable, we apply the index-1 condition for the coupled system (4.23), the implicit function theorem and Lemma 16 and get

$$\begin{aligned} \Delta z_{n+1} &= z_{n+1} - z(t_{n+1}) = G(w_{n+1}) - G(w(t_{n+1})) \\ \|\Delta z_{n+1}\| &= L_w^G \|\Delta w_{n+1}\|. \end{aligned} \tag{4.63}$$

To derive an estimation for the differential variable w, we investigate the stage value errors. Assuming exact values at t_n, i.e. $w_n = w(t_n),\ z_n = z(t_n)$, we have

$$\begin{aligned} W_1 - w(t_n) &= w_n + \tfrac{H}{2}\left(f(W_1,Z_1) - f(W_2,Z_2)\right) - w(t_n) \\ &= \tfrac{H}{2}\Big(f(W_1,Z_1) - f(W_2,Z_2)\Big) - \tfrac{H}{2}\Big(f(w(t_n),z(t_n)) - f(w(t_{n+1}),z(t_{n+1}))\Big) \\ &\quad + \tfrac{H}{2}\Big(f(w(t_n),z(t_n)) - f(w(t_{n+1}),z(t_{n+1}))\Big) \end{aligned}$$

For the second line we use a Taylor-series approximation and derive

$$\tfrac{H}{2}\Big(f(w(t_n),z(t_n)) - f(w(t_{n+1}),z(t_{n+1}))\Big) = \tfrac{H}{2}\Big(w'(t_n) - w'(t_{n+1}))\Big) = \mathscr{O}(H^2).$$

To the first line, we apply Lemma 16 and get

$$\|W_1 - w(t_n)\| = \tfrac{H}{2}\Big(\left(L_w^f + L_z^f L_w^G\right)\|W_1 - w(t_n)\| + \left(L_w^f + L_z^f L_w^G\right)\|W_2 - w(t_{n+1})\|\Big) + \mathscr{O}(H^2).$$

The stage value error at t_{n+1} is estimated in a similar way

$$\begin{aligned} W_2 - w(t_{n+1}) &= w_n + \tfrac{H}{2}\left(f(W_1,Z_1) + f(W_2,Z_2)\right) - w(t_{n+1}) \\ &= \tfrac{H}{2}\Big(f(W_1,Z_1) + f(W_2,Z_2)\Big) - \tfrac{H}{2}\Big(f(w(t_n),z(t_n)) + f(w(t_{n+1}),z(t_{n+1}))\Big) \\ &\quad + w_n - w(t_{n+1}) + \tfrac{H}{2}\Big(f(w(t_n),z(t_n)) + f(w(t_{n+1}),z(t_{n+1}))\Big) \end{aligned}$$

In the second line, we use Taylor expansion for $w(t_{n+1})$ and $f(w(t_{n+1}),z(t_{n+1}))$ and get

$$\begin{aligned} &w(t_n) - w(t_{n+1}) + \tfrac{H}{2}\Big(f(w(t_n),z(t_n)) + f(w(t_{n+1}),z(t_{n+1}))\Big) \\ &\qquad = -H\dot{w}(t_n) - \tfrac{H^2}{2}\ddot{w}(t_n) + \mathscr{O}(H^3) + \tfrac{H}{2}\left(2\dot{w}(t_n) + H\ddot{w}(t_n) + \mathscr{O}(H^2)\right) = \mathscr{O}(H^3) \end{aligned}$$

To the first line, we apply Lemma 16 an get the following estimation for the stage value error at t_{n+1}

$$\|W_2 - w(t_{n+1})\| = \tfrac{H}{2}\Big(\left(L_w^f + L_z^f L_w^G\right)\|W_1 - w(t_n)\| + \left(L_w^f + L_z^f L_w^G\right)\|W_2 - w(t_{n+1})\|\Big) + \mathscr{O}(H^3).$$

Combining both results, we end up with the following matrix inequality

$$\begin{pmatrix} 1 - \frac{H}{2}\left(L_w^f + L_z^f L_w^G\right) & -\frac{H}{2}\left(L_w^f + L_z^f L_w^G\right) \\ -\frac{H}{2}\left(L_w^f + L_z^f L_w^G\right) & 1 - \frac{H}{2}\left(L_w^f + L_z^f L_w^G\right) \end{pmatrix} \begin{pmatrix} \|W_1 - w(t_n)\| \\ \|W_2 - w(t_{n+1})\| \end{pmatrix} \leq \begin{pmatrix} \mathscr{O}(H^2) \\ \mathscr{O}(H^3) \end{pmatrix} \tag{4.64}$$

while the inequality has to be understood component wise. If the macro-step size is restricted to

$$H\left(L_w^f + L_z^f L_w^G\right) < 1$$

then the coefficient matrix in (4.64) is an M-matrix and in particular it is regular and its inverse is positive [Ple77]. We point out that (4.62) implies $W_2 - w(t_{n+1}) = \Delta w_{n+1}$. For simplicity of notation we set $C_1 := L_w^f + L_z^f L_w^G$ and get

$$\|\Delta w_{n+1}\| \leq \frac{1}{1-HC_1}\left(\tfrac{H}{2}C_1\mathscr{O}(H^2) + (1-\tfrac{H}{2}C_1)\mathscr{O}(H^3)\right) = \mathscr{O}(H^3). \tag{4.65}$$

Note that the macro-step in the Decoupled-Slowest-First approach coincides with the classical, single-rate LobattoIIIC integration method. LobattoIIIC fulfils the simplifying conditions $B(2), C(1), D(1)$. Applied to ODEs, LobattoIIIC has order of consistency 2 [But64]. For semi-explicit DAEs of index-1, LobattoIIIC preserves its order of consistency also for the algebraic variables since it is stiffly accurate and z_{n+1} does not depend on Z_1, cf. (4.63) or [BCP95].

Accuracy of the Fast Components:
We consider one micro-step $t_n + (l-1)h \to t_n + lh$ as described in (4.56-4.57). We introduce the short-hand notation $t_{n+(l-1)/m} = t_n + (l-1)h$ and $t_{n+l/m} = t_n + lh$. For the error in the fast changing, algebraic variable we have

$$\Delta z_{Fn+l/m} = Z_{F2}^l - z_F(t_{n+l/m}) = G_F(W_{F2}^l, \bar{W}_{S2}^l) - G(w_F(t_{n+l/m}), w_S(t_{n+l/m})).$$

Lemma 16 and Lemma 18 give

$$\|\Delta z_{Fn+l/m}\| \leq L_{w_F}^{G_F}\|W_{F2}^l - w_F(t_{n+l/m})\| + C_w^l\|\Delta w_{Sn+1}\| + \mathscr{O}(H^2)$$

for a constant $C_w^l > 0$.

For the error in the fast changing, differential subsystem $\Delta w_{Fn+l/m}$, we consider the stage value errors $W_{F1}^l - w_F(t_{n+(l-1)/m})$ and $W_{F2}^l - w_F(t_{n+l/m})$ and proceed as for the macro-step. We derive the following estimates

$$\begin{aligned}
\|W_{F1}^l - w_F(t_{n+(l-1)/m})\| \leq \|\Delta w_{Fn+(l-1)/m}\| + \tfrac{h}{2}\Big\{ & (L_{w_F}^{f_F} + L_{z_F}^{f_F}L_{w_F}^{G_F})\|W_{F1}^l - w_F(t_{n+(l-1)/m})\| \\
& + (L_{w_F}^{f_F} + L_{z_F}^{f_F}L_{w_F}^{G_F})\|W_{F2}^l - w_F(t_{n+l/m})\| \\
& + C_w^l\|\Delta w_{Sn+1}\| + C_z^l\|\Delta z_{Sn+1}\| + \mathscr{O}(H^2)\Big\} + \mathscr{O}(h^2) \\
\|W_{F2}^l - w_F(t_{n+l/m})\| \leq \|\Delta w_{Fn+(l-1)/m}\| + \tfrac{h}{2}\Big\{ & (L_{w_F}^{f_F} + L_{z_F}^{f_F}L_{w_F}^{G_F})\|W_{F1}^l - w_F(t_{n+(l-1)/m})\| \\
& + (L_{w_F}^{f_F} + L_{z_F}^{f_F}L_{w_F}^{G_F})\|W_{F2}^l - w_F(t_{n+l/m})\| \\
& + C_w^l\|\Delta w_{Sn+1}\| + C_z^l\|\Delta z_{Sn+1}\| + \mathscr{O}(H^2)\Big\} + \mathscr{O}(h^3)
\end{aligned}$$

for a constant $C_z^l > 0$. For Δw_{Sn+1}, Δw_{Sn+1} we insert the results of the analysis of the macro-step (4.63, 4.65). We set $C_2 := L_{w_F}^{f_F} + L_{z_F}^{f_F}L_{w_F}^{G_F}$ and get the following (component wise) matrix inequality

$$\begin{pmatrix} 1-\frac{h}{2}C_2 & -\frac{h}{2}C_2 \\ -\frac{h}{2}C_2 & 1-\frac{h}{2}C_2 \end{pmatrix}\begin{pmatrix} \|W_{F1}^l - w_f(t_{n+(l-1)/m})\| \\ \|W_{F2}^l - w_f(t_{n+l/m})\| \end{pmatrix} \leq \begin{pmatrix} \|\Delta w_{Fn+(l-1)/m}\| + \mathscr{O}(H^2) \\ \|\Delta w_{Fn+(l-1)/m}\| + \mathscr{O}(H^3) \end{pmatrix}. \tag{4.66}$$

Note that the fixed multirate factor m implies $\mathscr{O}(h^n) = \mathscr{O}(H^n)$ for $n \in \{2,3\}$. The M-Matrix

condition for coefficient matrix in (4.66) leads to the restriction for the micro-step size

$$h(L_{w_F}^{f_F} + L_{z_F}^{f_F} L_{w_F}^{G_F}) < 1$$

that guarantees existence and positivity of the inverse of the coefficient matrix and we end up with

$$\|\Delta w_{Fn+l/m}\| \leq \frac{1}{1-hC_2}\left(\|\Delta w_{Fn+(l-1)/m}\| + \mathscr{O}(H^3)\right). \tag{4.67}$$

Summing up over $l = 1,\ldots,m$ and using exact values at t_n (4.59) complete the proof. □

After proving the order of consistency of the mrIRK2-DAE scheme, it is now to show its convergence.

4.3.3 Convergence of mrIRK2-DAE

After proving consistency of the mrIRK2-DAE scheme, it is now to show that the scheme is converging for a fixed macro-step size H. We proceed as for the mrIRK1-DAE scheme in Section 4.3.3: For any variable x of the DAE-IVP (4.18-4.21), let

$$E(x,t_n) = x_n - x(t_n)$$

denote the global error between the numerical approximation x_n at time t_n computed by the mrIRK2-DAE scheme and the analytical solution $x(t_n)$. We show, that the update function for the algebraic variables of the mrIRK2-DAE scheme fulfils the conditions of the convergence theorem in [DHZ87]. The following Proposition works out the details.

Proposition 27. *We consider the index-*1 *DAE-IVP (4.18-4.21) fulfilling Assumption 15. We apply the mrIRK2-DAE method (4.54-4.55) and (4.56-4.57). The macro-step size H and the multirate facor m are chosen according to the step size restrictions* (4.60) *such that the method is consistent of order* 2*. Then the mrIRK2-DAE method converges and we get for the global error*

$$E(w_F,t_n) = \mathscr{O}(H^2), \quad E(z_F,t_n) = \mathscr{O}(H^2), \quad E(w_S,t_n) = \mathscr{O}(H^2), \quad E(z_S,t_n) = \mathscr{O}(H^2).$$

Proof. Analogously to the proof of Theorem 25, the assumptions of the Theorem in [DHZ87] have to be checked:

One-Step Method. As for the mrIRK1-DAE scheme, the mrIRK2-DAE scheme computes the approximations $w_{Fn+1}, z_{Fn+1}, w_{Sn+1}, z_{Sn+1}$ at t_{n+1} as functions of the approximations $w_{Fn}, z_{Fn}, w_{Sn}, z_{Sn}$ at t_n.

Consistency. Proposition 26 guarantees order of consistency 2 for w_F, w_S, z_S and 1 for z_F if the step sizes are chosen according to (4.60).

Perturbation Condition (4.48). The Perturbation condition holds for all stiffly-accurate, k-stage Runge-Kutta methods with coefficient matrix A, stage-vector c and weight vector b if the Coupled-Slowest-First-Approach is applied:

The approximation of the slow algebraic variable z_S at $t_n + H$ is given by

$$\begin{aligned} z_{Sn+1} &= G_S(w^*_{Fn+1}, w_{Sn+1}) \\ &= G_S\Big(w_{Fn} + H\sum_{i=1}^{k} b_i f_F(W^*_{Fi}, Z^*_{Fi}, W_{Si}, Z_{Si}),\ w_{Sn} + H\sum_{i=1}^{k} b_i f_S(W^*_{Fi}, Z^*_{Fi}, W_{Si}, Z_{Si})\Big) \end{aligned}$$

For the fast algebraic variables we have

$$\begin{aligned} z_{Fn+1} &= G_F(w_{Fn+1}, w_{Sn+1}) \\ &= G_F\Big(w_{Fn} + \tfrac{H}{m}\sum_{i=1}^{k} b_i \sum_{l=1}^{m} f_F(W^l_{Fi}, Z^l_{Fi}, \bar{W}^l_{Si}, \bar{Z}^l_{Si}), w_{Sn} + H\sum_{i=1}^{k} b_i f_S(W^*_{Fi}, Z^*_{Fi}, W_{Si}, Z_{Si})\Big) \end{aligned}$$

Next, we compute

$$\frac{\partial \Psi_S}{\partial z_n}, \qquad \text{and} \qquad \frac{\partial \Psi_F}{\partial z_n}$$

with Ψ_S the update function of the slow algebraic variables, Ψ_F the update function of the fast algebraic variables and $z_n = (z_{Fn}^\top, z_{Sn}^\top)^\top$. Evaluating both expressions for $H = 0$, we finally get

$$\left.\frac{\partial \Psi_S}{\partial z_n}\right|_{H=0} = 0, \qquad \text{and} \qquad \left.\frac{\partial \Psi_F}{\partial z_n}\right|_{H=0} = 0$$

and the proof of Theorem 25 is complete.

□

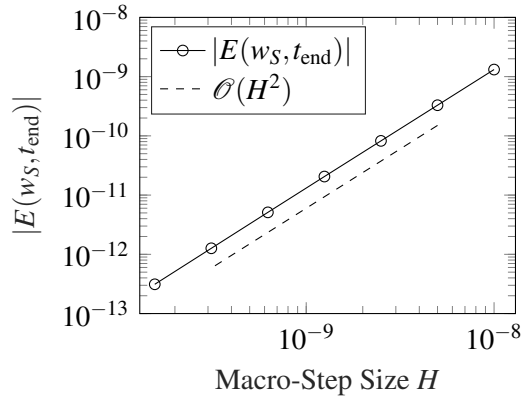
After the analytical proof the convergence of the mrIRK2-DAE scheme, the following section provides a numerical example that illustrates the results from above.

4.3.4 Numerical Results

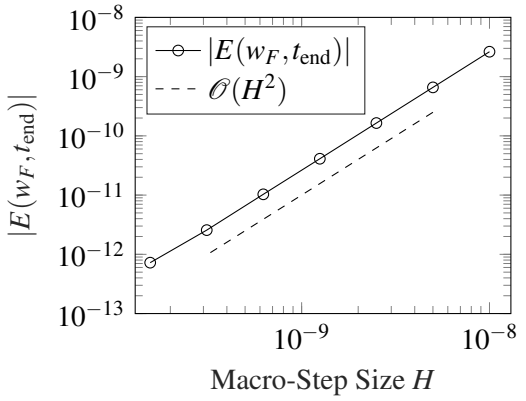
We apply the mrIRK2-DAE scheme to the extended Prothero-Robinson equation (4.49) with parameters (4.50) and initial values (4.51). We integrate the DAE-IVP over the time interval $[t_0, t_{\text{end}}]$ with macro-step size $2^{-i} \cdot 10^{-8}$, $i = 0, 1, \ldots, 6$ and fixed multirate factor $m = 10$. Figure 4.4 shows the absolute value of the global error at t_{end} with regard to the macro-step size for each component. We observe order 2 convergence for all variables.

Section Summary

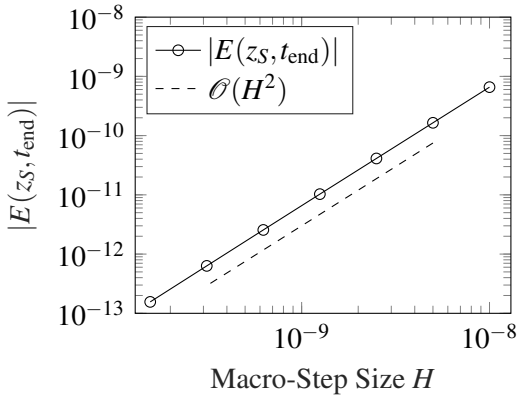
We established the mrIRK2-DAE scheme: a multirate time integration method for semi-explicit DAEs of index-1 based on the LobattoIIIC-scheme. The coupling between the subsystems of different dynamics was realised by the Coupled-Slowest-First approach for the macro-step and linear interpolation for the micro-steps. We showed consistency and convergence of the mrIRK2-scheme and derived the order of convergence 2. A numerical simulation confirmed the previous theoretical results. We saw, that a multirate scheme for semi-explicit DAEs of index-1 with this particular coupling strategy has at most convergence order 2 due to the linear interpolation of the slow changing variables.



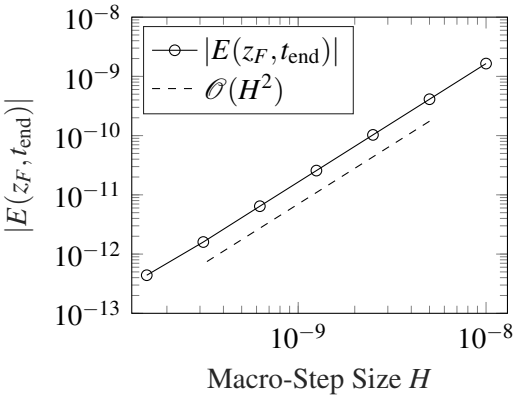
(a) $E(w_s,t_{\text{end}})$ in $\mathscr{O}(H^2)$



(b) $E(w_F,t_{\text{end}})$ in $\mathscr{O}(H^2)$



(c) $E(z_S,t_{\text{end}})$ in $\mathscr{O}(H^2)$



(d) $E(z_F,t_{\text{end}})$ in $\mathscr{O}(H^2)$

Figure 4.4: Order of convergence for the mrIRK2-DAE scheme ($m = 10$): a)-d) order 2.

4.4 Decoupled Multirate One-Step Methods – A Link to Dynamic Iteration

In the previous chapters we investigated multirate methods for semi-explicit DAEs of index-1 that were based on a particular Runge-Kutta method (implicit Euler and LobattoIIIC). Consistency and convergence have been proven for both integration methods. In the following we provide a convergence theorem for the class of all Decoupled-Slowest-First multirate one-step methods where the basic integration scheme is of order p and the error of the inter- and extrapolated values is at least of order $p-1$. To this end, we interpret the multirate time integration method as a dynamic iteration scheme. We define the splitting functions for the multirate method and derive assumptions to the underlying DAE-system such that convergence of the method can be guaranteed.

We briefly discuss the basic concept of **dynamic iteration** and introduce the necessary notation.

We consider the coupled DAE-IVP (4.18-4.21) in the compact formulation of (4.61) and compute an approximation on the interval $[t_0,t_{\text{end}}]$ using the dynamic iteration scheme. Using this scheme, continuous approximations can be achieved. Therefore, an approximation on the interval $(t_i,t_{i+1}]$ of an analytical solution $x(t)$ is denoted by the waveform $\tilde{x}|_{(t_i,t_{i+1}]}(t)$.

Before the dynamic iteration scheme starts, the time interval $[t_0, t_{\text{end}}]$ is split into N **time windows** $[t_n, t_{n+1}]$ for $n \in \{0, \ldots, N-1\}$ with

$$t_0 < t_1 < \ldots < t_{N-1} < t_N = t_{\text{end}}.$$

The dynamic iteration scheme computes the approximations by an iterative process on each time window. We assume that the approximation on the time window $[t_{n-1}, t_n]$ is available. The initialisation of the iterative process is done by an extrapolation of the data of the previous time window to the current time window

$$\begin{pmatrix} \tilde{w}^0|_{(t_n,t_{n+1}]} \\ \tilde{z}^0|_{(t_n,t_{n+1}]} \end{pmatrix} = \check{\Phi} \begin{pmatrix} \tilde{w}^k|_{(t_{n-1},t_n]}(t_n) \\ \tilde{z}^k|_{(t_{n-1},t_n]}(t_n) \end{pmatrix}. \tag{4.68}$$

with $\tilde{w}|_{(t_n,t_{n+1}]}$, $\tilde{w}|_{(t_n,t_{n+1}]}$ the waveform of the approximation of the differential and algebraic subsystem, respectively. The extrapolation is realised by the operator $\check{\Phi}$ and we assume that k iterations have been carried out on the time window $[t_{n-1}, t_n]$.

The dynamic iteration scheme is represented by the iteration operator

$$\check{\Psi}: \quad \begin{pmatrix} \tilde{w}^{k-1}|_{(t_n,t_{n+1}]} \\ \tilde{z}^{k-1}|_{(t_n,t_{n+1}]} \end{pmatrix} \to \begin{pmatrix} \tilde{w}^k|_{(t_n,t_{n+1}]} \\ \tilde{z}^k|_{(t_n,t_{n+1}]} \end{pmatrix}. \tag{4.69}$$

If the scheme is converging, (4.69) is performed until a given accuracy is reached. For coupled systems of ODEs, a dynamic iteration scheme is always convergent [Bur95]. For coupled systems of DAEs, additional assumptions to the DAE system and the iteration scheme have to be fulfilled to guarantee convergence of the scheme [AG01].

The properties of a specific dynamic iteration scheme are defined by its **splitting functions** $\check{F}$ and $\check{G}$. The splitting functions describes the dependencies between the subsystems and the iteration steps. The general splitting functions for the subsystems of the DAE-IVP (4.18-4.21) in the compact formulation (4.61) read [Bar04]:

$$\dot{\tilde{w}}_i = \check{F}_i\left(\tilde{w}|^k_{(t_n,t_{n+1}]}, \tilde{w}|^{k-1}_{(t_n,t_{n+1}]}, \tilde{z}|^k_{(t_n,t_{n+1}]}, \tilde{z}|^{k-1}_{(t_n,t_{n+1}]}\right) \tag{4.70}$$

$$0 = \check{G}_i\left(\tilde{w}|^k_{(t_n,t_{n+1}]}, \tilde{w}|^{k-1}_{(t_n,t_{n+1}]}, \tilde{z}|^k_{(t_n,t_{n+1}]}, \tilde{z}|^{k-1}_{(t_n,t_{n+1}]}\right) \tag{4.71}$$

for $i \in \{F, S\}$. The initial values are given by $\tilde{w}^k_n(t_n) = \tilde{w}^0_n(t_n)$ and $\tilde{z}^k_n(t_n) = \tilde{z}^0_n(t_n)$ for all k. Setting $\check{F} := (\check{F}_F^\top, \check{F}_S^\top)^\top$ and $\check{G} := (\check{G}_F^\top, \check{G}_S^\top)^\top$ we derive the splitting functions for the coupled DAE-IVP (4.18-4.21). The splitting functions have to fulfil the compatibility conditions [AG01]

$$\check{F}(w,w,z,z) = f(w,z) \qquad \text{and} \qquad \check{G}(w,w,z,z) = g(w,z) \tag{4.72}$$

with w, z the analytical solutions of (4.18-4.21) in the compact formulation of (4.61) and f, g given in (4.61).

A multirate time integration method that is based on a one-step method can be interpreted as a dynamic iteration scheme: The time windows correspond to the macro-steps and on each time window one iteration is carried out ($k = 1$). This interpretation enables us to derive analytical results for multirate one-step methods in a general formulation. The following Theorem provides a convergence proof for Decoupled-Slowest-First multirate one-step methods based on the contractivity condition in [AG01].

Theorem 28. *We consider the coupled DAE-IVP (4.18-4.21) with fulfilling Assumption 15 and the index-*1 *condition for the subsystems* (4.22) *and the coupled system* (4.23)*. We apply a multirate one-step method using the Decoupled-Slowest-First approach and integrate the DAE-IVP over the time interval* $[t_0, t_{end}]$*. If*

(i) the basic integration scheme for the each subsystem is of order p in the single-rate case,

(ii) the error of the extrapolated and interpolated coupling terms is at least of order $p-1$*,*

(iii) the algebraic variables are always consistently computed,

(iv) the following conditions are fulfilled

$$\alpha_1 := \frac{L_{z_F}^{g_S}}{L_{z_S}^{g_S}} < \frac{1}{L} \qquad \text{and} \qquad \alpha_2 := \frac{L_{z_S}^{g_F}}{L_{z_F}^{g_F}} < 1 \tag{4.73}$$

with L the Lipschitz-constant of the extrapolation operator $\check{\Phi}$ (4.68)*,*

then the time integration is stable and the multirate method has still order p.

Proof. We follow the proof of Theorem 2.2 in [AG01] in the more general problem setting of [BBGS13]. We interpret the multirate one-step method as a dynamic iteration scheme with one iteration $k = 1$.

The proof is organised as follows: In *a)* we define the splitting function of the multirate method and introduce the necessary notation. In part *b)* we show that the integration method is stable if a contractivity condition is fulfilled. In *c)* we deduce the inequalities in (4.73) from the contractivity condition in *b)* if the global error of the integration method is uniformly bounded. The proof of boundedness of the global error in *d)* concludes the argumentation.

a) The splitting function for a multirate one-step method with Decoupled-Slowest-First coupling approach applied to the DAE-IVP (4.18-4.21) reads

$$\begin{aligned} \dot{w}^k = \begin{pmatrix} \dot{w}_F^k \\ \dot{w}_S^k \end{pmatrix} = \check{F}(w_F^k, w_S^k, w_F^{k-1}, z_F^k, z_S^k, z_F^{k-1}) &= \begin{pmatrix} \check{F}_F(w_F^k, w_S^k, z_F^k, z_S^k) \\ \check{F}_S(w_S^k, w_F^{k-1}, z_S^k, z_F^{k-1}) \end{pmatrix} \\ 0 = \check{G}(w_F^k, w_S^k, w_F^{k-1}, z_F^k, z_S^k, z_F^{k-1}) &= \begin{pmatrix} \check{G}_F(w_F^k, w_S^k, z_F^k, z_S^k) \\ \check{G}_S(w_S^k, w_F^{k-1}, z_S^k, z_F^{k-1}) \end{pmatrix} \end{aligned}. \tag{4.74}$$

We consider the time window $[t_n, t_{n+1}]$ with $H = t_n - t_{n-1}$ and introduce the short-hand notation

$$w_n = w|_{(t_n, t_{n+1}]}, \qquad z_n = z|_{(t_n, t_{n+1}]}.$$

for the waveform of the analytical solution of (4.18-4.21) in the compact formulation of (4.61). The waveform of the numerical approximation is denoted by $(\tilde{w}_n^k, \tilde{z}_n^k)$ with iteration index k.

The Decoupled-Slowest-First multirate method extrapolates the values of the fast subsystem to compute an approximation of the slow subsystem. Therefore, the splitting function for the slow subsystem $\check{F}_S, \check{G}_S$ depend on old iterates of the fast subsystem w_F^{k-1}, z_F^{k-1}. We

define the global error

$$\begin{pmatrix} \varepsilon_{wn} \\ \varepsilon_{zn} \end{pmatrix} := \begin{pmatrix} \tilde{w} - w \\ \tilde{z} - z \end{pmatrix}\Bigg|_{(t_n, t_{n+1}]} = (\breve{\Psi} \circ \breve{\Phi}) \begin{pmatrix} \tilde{w}^1_{n-1}(t_n) \\ \tilde{z}^1_{n-1}(t_n) \end{pmatrix} - \breve{\Psi} \begin{pmatrix} w_n \\ z_n \end{pmatrix}$$

with the extrapolation operator $\breve{\Phi}$, see (4.68) and the iteration operator of the multirate scheme $\breve{\Psi}$ as described in (4.69). The considered multirate time methods are one-step integration schemes. Therefore, only the values at t_n can be used for the extrapolation of the variables on the time window $[t_n, t_{n+1}]$. We point out that the analytical solution (w_n, z_n) is a fix point of the mapping $\breve{\Psi}_n$. We assume, that the error is bounded by a constant $\beta > 0$

$$\|\varepsilon_{wm}\| + \|\varepsilon_{zm}\| \leq \beta \tag{4.75}$$

for all $0 \leq m \leq n$ with $t_n \leq t_{\text{end}}$. All error estimates in this proof are given in the L^∞-norm with respect to the considered time window:

$$\|\varepsilon\| := \|\varepsilon\|_\infty = \inf\left\{D \geq 0 : \ \|\varepsilon(x)\| \leq D \quad \forall\, x \in [t_n, t_{n+1}]\right\}.$$

b) For a detailed analysis of the global error, we split ε_n into

$$\begin{pmatrix} \varepsilon_{wn} \\ \varepsilon_{zn} \end{pmatrix} = \underbrace{(\breve{\Psi} \circ \breve{\Phi}) \begin{pmatrix} \tilde{w}^1_{n-1}(t_n) \\ \tilde{z}^1_{n-1}(t_n) \end{pmatrix} - (\breve{\Psi} \circ \breve{\Phi}) \begin{pmatrix} w(t_n) \\ z(t_n) \end{pmatrix}}_{=:E_n} + \underbrace{(\breve{\Psi} \circ \breve{\Phi}) \begin{pmatrix} w(t_n) \\ z(t_n) \end{pmatrix} - \breve{\Psi} \begin{pmatrix} w_n \\ z_n \end{pmatrix}}_{=:\Delta_n}$$

with $E_n = (E_{wn}^\top, E_{zn}^\top)^\top$ the error propagation from the previous time window and $\Delta_n = (\Delta_{wn}^\top, \Delta_{zn}^\top)$ the error of the current time window. For error estimates for E_n and Δ_n, we follow the lines of the proof of Theorem 2.2 in [AG01] for the case of $k = 1$.

- For E_n, we apply Lemma 3.2 in [AG01] and get the error recursion

$$\begin{pmatrix} \|E_{wn}\| \\ \|E_{zn}\| \end{pmatrix} \leq \begin{pmatrix} 1 + D_1^* H & D_1^* H \\ D_1^* & \alpha_n^* \end{pmatrix} \cdot \begin{pmatrix} \|E_{wn-1}\| \\ \|E_{zn-1}\| \end{pmatrix}, \tag{4.76}$$

with a constant $D_1^* > 0$ and

$$\alpha_n^* = L\left(\hat{\alpha} + \frac{4DH}{\frac{\hat{\alpha}}{2D} + \sqrt{H}}\right)$$

for $D > 0$ and L the Lipschitz constant of the extrapolation operator $\breve{\Phi}$. The parameter $\hat{\alpha}$ is given by

$$\begin{aligned} \hat{\alpha} = \alpha + \mathscr{O}(1)\Big(&\left\|\breve{\Phi}\left(\tilde{w}^1_{n-1}(t_n)\right) - w_n\right\| + \left\|\breve{\Phi}\left(\tilde{z}^1_{n-1}(t_n)\right) - z_n\right\| \\ &+ \left\|\breve{\Phi}\left(w(t_n)\right) - w_n\right\| + \left\|\breve{\Phi}\left(z(t_n)\right) - z_n\right\| \Big) \end{aligned} \tag{4.77}$$

with

$$\alpha = \left\| \left(\frac{\partial \breve{G}}{\partial z^k}\right)^{-1} \frac{\partial \breve{G}}{\partial z^{k-1}} \right\| \tag{4.78}$$

cf. [BBGS13].

- To derive an estimate for the local error Δ_n, we introduce the extrapolation error $\check{\delta}_n = \|\check{\delta}_{wn}\| + \|\check{\delta}_{zn}\|$ and we have
$$\check{\delta}_n = \mathscr{O}(H^p).$$
Once again, we apply Lemma 3.2 in [AG01] and estimate
$$\|\Delta_{wn}\| + H\|\Delta_{zn}\| \leq D_2^* H \check{\delta}_n \tag{4.79}$$
with a constant D_2^* that is independent of H and $\|\Delta_{wn}\| = \mathscr{O}(H^{p+1})$, $\|\Delta_{zn}\| = \mathscr{O}(H^p)$ by assumption.

- Combining the results of (4.76) and (4.79) gives
$$\begin{pmatrix} \|\varepsilon_{wn}\| \\ \|\varepsilon_{zn}\| \end{pmatrix} \leq \begin{pmatrix} 1 + D_1^* H & D_1^* H \\ D_1^* & \alpha^* \end{pmatrix} \cdot \begin{pmatrix} \|E_{wn-1}\| \\ \|E_{zn-1}\| \end{pmatrix} + \begin{pmatrix} D_2^* H \check{\delta}_n \\ D_2^* \check{\delta}_n \end{pmatrix} \tag{4.80}$$
for all time windows $[t_n, t_{n+1}]$ and $n \geq 0$ with $t_n \leq t_{\text{end}}$. We set $\alpha^* = \max_{m \leq n} \alpha_m^*$ and $\varepsilon_{w,-1} = \varepsilon_{z,-1} := 0$. For $\alpha^* < 1$, the error recursion (4.80) is stable [DHZ87] and we have
$$\|\varepsilon_{wn}\| + \|\varepsilon_{zn}\| \leq D^* \max_{0 \leq m < n} \check{\delta}_n \tag{4.81}$$
with a constant $D^* > 0$, which does not depend on H and n. This proves the stability of the multirate time integration method if the contractivity condition is fulfilled. Then, the order of the integration scheme is given by the applied extrapolation method.

c) We show that the contractivity condition $\alpha^* < 1$ can be guaranteed if (4.73) holds. For α^* we have
$$\alpha^* \geq \alpha_n^* = L\left(\hat{\alpha} + \frac{4DH}{\frac{\hat{\alpha}}{2D} + \sqrt{H}}\right) = L\left(\hat{\alpha} + \mathscr{O}(\sqrt{H})\right) \tag{4.82}$$
with $\hat{\alpha}$ defined in (4.77). We estimate the right-hand side of (4.77): For the first summand, we derive
$$\left\|\check{\Phi}_n(w(t_n)) - w_n\right\| + \left\|\check{\Phi}_n(z(t_n)) - z_n\right\| = \check{\delta}_n = \mathscr{O}(H^p).$$
Using assumption (4.75), we get for for the second summand
$$\begin{aligned}
&\left\|\check{\Phi}_n\left(\tilde{w}_{n-1}^1(t_n)\right) - w_n\right\| + \left\|\check{\Phi}_n\left(\tilde{z}_{n-1}^1(t_n)\right) - z_n\right\| \\
&\qquad = \left\|\check{\Phi}_n\left(\tilde{w}_{n-1}^1(t_n)\right) - \check{\Phi}_n\left(w_{n-1}(t_n)\right) + \check{\Phi}_n\left(w_{n-1}(t_n)\right) - w_n\right\| \\
&\qquad\quad + \left\|\check{\Phi}_n\left(\tilde{z}_{n-1}^1(t_n)\right) - \check{\Phi}_n\left(z_{n-1}(t_n)\right) + \check{\Phi}_n\left(z_{n-1}(t_n)\right) - z_n\right\| \\
&\qquad \leq \left\|\check{\Phi}_n\left(\tilde{w}_{n-1}^1(t_n)\right) - \check{\Phi}_n\left(w_{n-1}(t_n)\right)\right\| + \left\|\check{\Phi}_n\left(w_{n-1}(t_n)\right) - w_n\right\| \\
&\qquad\quad + \left\|\check{\Phi}_n\left(\tilde{z}_{n-1}^1(t_n)\right) - \check{\Phi}_n\left(z_{n-1}(t_n)\right)\right\| + \left\|\check{\Phi}_n\left(z_{n-1}(t_n)\right) - z_n\right\| \\
&\qquad \leq L\left\|\tilde{w}_{n-1}^1(t_n) - w_{n-1}(t_n)\right\| + \mathscr{O}(H^p) \\
&\qquad\quad + L\left\|\tilde{z}_{n-1}^1(t_n) - z_{n-1}(t_n)\right\| + \mathscr{O}(H^p) \\
&\qquad \leq L\left(\|\varepsilon_{wn-1}\| + \|\varepsilon_{zn-1}\|\right) + \mathscr{O}(H^p) \\
&\qquad = \mathscr{O}(\beta) + \mathscr{O}(H^p).
\end{aligned}$$
Adding both estimations, we end up with
$$\hat{\alpha} = \alpha + \mathscr{O}(\beta) + \mathscr{O}(H^p)$$

and $\hat{\alpha} < 1$ can always be guaranteed by choosing β and H sufficiently small. For the splitting functions (4.74), we compute α according to (4.78)

$$\frac{\partial \check{G}}{\partial z^k} = \begin{pmatrix} \frac{\partial g_S}{\partial z_S} & 0 \\ \frac{\partial g_F}{\partial z_S} & \frac{\partial g_F}{\partial z_F} \end{pmatrix}, \qquad \frac{\partial \check{G}}{\partial z^{k-1}} = \begin{pmatrix} 0 & \frac{\partial g_S}{\partial z_F} \\ 0 & 0 \end{pmatrix},$$

with (4.77) and (4.82) we derive the contractivity condition

$$\alpha = \left\| \left(\frac{\partial \check{G}}{\partial z^k} \right)^{-1} \frac{\partial \check{G}}{\partial z^{k-1}} \right\| = \left\| \begin{pmatrix} 0 & \left(\frac{\partial g_S}{\partial z_S}\right)^{-1} \frac{\partial g_S}{\partial z_F} \\ 0 & \left(\frac{\partial g_F}{\partial z_F}\right)^{-1} \frac{\partial g_F}{\partial z_S} \left(\frac{\partial g_S}{\partial z_S}\right)^{-1} \frac{\partial g_S}{\partial z_F} \end{pmatrix} \right\| < \frac{1}{L}.$$

This condition is fulfilled, if

$$\left\| \left(\frac{\partial g_S}{\partial z_S}\right)^{-1} \frac{\partial g_S}{\partial z_F} \right\| < \tfrac{1}{L} \quad \text{and} \quad \left\| \left(\frac{\partial g_F}{\partial z_F}\right)^{-1} \frac{\partial g_F}{\partial z_S} \left(\frac{\partial g_S}{\partial z_S}\right)^{-1} \frac{\partial g_S}{\partial z_F} \right\| < \tfrac{1}{L}$$

holds [BG20]. Using the notation of (4.73), we get

$$\alpha_1 < \tfrac{1}{L} \qquad \text{and} \qquad \alpha_1 \alpha_2 < \tfrac{1}{L}$$

and end up with an equivalent formulation of the contractivity condition (4.73).

d) It remains to show that the assumption (4.75) holds if H is chosen sufficiently small. The boundedness of $\|\varepsilon_{wm}\| + \|\varepsilon_{zm}\|$ is a direct consequence of (4.81) and $\check{\delta}_n = \mathscr{O}(H^p)$. We sketch an induction over all time windows m with $0 \leq m \leq n$ and $t_n \leq t_{\text{end}}$.

- For $m = 0$ the statement is obviously correct since the initial values are chosen consistently.
- If the statement is correct for $m = n-1$ and $\check{\delta}_n > \check{\delta}_p$ for all $p < n$ we set $H^* < H$ such that $\|\varepsilon_n^*\| \leq \beta$ while ε_n^* denotes the global error between the analytical solution and a numerical approximation computed with step size H^*. This can always be done since $\check{\delta}_n = \mathscr{O}(H^p)$.

□

In Section 4.2.3 and 4.2.4 we showed consistency and convergence for the mrIRK1-DAE scheme using the Decoupled-Slowest-First approach. Here, consistency and convergence could be guaranteed by choosing a sufficiently small macro-step size H and an inherent multirate factor m. The assumptions of (4.73) in Theorem 28 are even stricter and convergence can only be guaranteed for a class of DAE-IVPs that fulfil these assumptions. A closer look at the contractivity condition (4.73) leads to the necessity of a weak coupling between the algebraic subsystems. A DAE-IVP that suits well for multirate time integration usually fulfils such a weakly coupled structure, at least for the fast-to-slow coupling (2.6).

If the contractivity condition (4.73) is violated, but we have $\alpha < 1$ in (4.78), it is possible to overcome this problem and guarantee a stable time integration by carrying out more iterations $k > 1$ on each time-step. The resulting time integration scheme is not a multirate one-step method in the classical sense but this strategy leads to larger class of iterative multirate methods.

In Section 4.2, we did a detailed convergence analysis of the mrIRK1-DAE scheme using the

Decoupled-Slowest-First approach without using the techniques and results of dynamic iteration methods. The results of the convergence analysis of the current section are more general and can be applied to *any* multirate one-step method using the Decoupled-Slowest-First approach. However, the more general approach leads to assumptions that have to be fulfilled by the considered DAE-system, i.e. if the DAE-system violates the contractivity condition (4.73), the theorem cannot guarantee the convergence of the integration scheme. In Section 4.2 we showed, the we can guarantee the convergence of the mrIRK1-scheme by adapting the integration parameters H and m, even if the Lipschitz constants of the considered DAE-system are large.

Chapter Summary

To derive a multirate one-step method for semi-explicit DAEs of index-1, we started with classical single-rate Runge-Kutta schemea which are implicit and stiffly-accurate: the implicit Euler method and the LobattoIIIC scheme. Both methods have been extended to a multirate integration scheme using different step sizes according to the dynamical behaviour of the subsystems of the coupled system of DAEs. For the multirate implicit Euler method, we implemented all three established coupling approaches: Decoupled-Slowest-First, Coupled-Slowest-First and Coupled-First-Step. We showed, that the Coupled-Slowest-First approach leads to a higher order of consistency compared to the other approaches. Therefore, we used only this approach for the multirate LobattoIIIC method. We showed analytically and numerically convergence order 1 for the multirate implicit Euler method and order 2 for the multirate LobattoIIIC scheme. The convergence can be guaranteed by adapting the macro- and the micro-step size according to the properties of the considered DAE-IVP. The main result of this section can be summarised as follows: The convergence order of the underlying single-rate scheme can be maintained for the derived multirate method, if the coupling variables are evaluated with a sufficient accuracy. Due to a linear interpolation during the integration of the fast changing subsystem, the convergence order of the resulting multirate scheme is bounded by 2. To derive a multirate method of higher order in a similar war, a more accurate strategy to evaluate the coupling variables has to be developed.

Finally, we linked the theories of multirate time integration and dynamic iteration. By interpreting a multirate method as a dynamic iteration scheme, we derived a convergence theorem for a general multirate one-step method using the Decoupled-Slowest-First approach. However, the convergence of the multirate integration scheme depends in this case on the properties of the considered DAE-IVP.

5 Chapter 5
Summary

Multirate methods for an efficient time integration of multiscale differential equations have been discussed in research and software development for several decades. In this thesis, we discussed multirate methods for two special types of multiscale differential equations:

The **first part** of the thesis dealt with multirate methods for multiscale ordinary differential equations (ODEs) with a high dimensional, linear-affine slow subsystem. Applying a model order reduction (MOR) to the slow changing subsystem, it is projected onto a low dimensional replacement system. Using a multirate method for time integration, the number of function evaluations of the slow subsystem decreases significantly, the MOR leads to a smaller dimension of the slow subsystem and we expect an additional gain of efficiency. This leads to the first important result of this thesis:
1. We showed that the MOR of the slow subsystem only results in a shorter computation time if the coupling interface of the fast subsystem to the slow subsystem is of small dimension. Especially for implicit integration methods, a high dimensional coupling interface leads to high dimensional Jacobian-matrices and the computation time of the multirate method does not decrease as expected. Since the fast subsystem does not depend on the detailed information of every single slow component, a low dimensional coupling interface can be defined for many multiscale problems. This can be done for example by exploiting the underlying physical properties of the coupled system. Our simulation results showed, that a MOR of the slow subsystem and a small dimensional coupling interface decrease the computation time of the applied multirate integration scheme significantly.
2. Beside the gain of efficiency by the MOR, we investigated the approximation properties of the coupled multiscale ODE with an order reduced, slow subsystem. For a coupled, linear-affine multiscale ODE, we derived a combined error bound in time domain, which estimates the MOR caused error and the integration error of the multirate method. Both errors can be estimated separately. The MOR caused error describes an error in the mathematical modelling of the system, so a multirate time integration only makes sense, if the MOR-caused error is sufficiently small. Then, the parameters of the multirate method are chosen according to the properties of the coupled multiscale ODE with order reduced, slow subsystem to derive efficiently a reliable approximation of the dynamical behaviour of the coupled multiscale ODE.

References to current work and outlook:
In this work, we consider multiscale ODEs with a linear-affine slow subsystem. In many applications, we can assume that a linearisation can be applied to the slow subsystem if non-linearities occur. If this is not the case or a linearisation leads to large errors, methods of non-linear MOR have to be applied to project the slow subsystem onto a low dimensional replacement system. An error bound for the widely used non-linear MOR technique proper orthogonal decomposition (POD) is given in [CS12]. For the particular case of coupled electrical circuits, there are results for non-linear MOR of coupled systems [SS13]. The combination of multirate time integration and MOR for non-linear systems is investigated in [BCG20]. For the derivation of the combined error bound in the linear setting, we used a close relation between balanced truncation MOR and POD applied to linear systems. This close relation can be a starting point for further investigations

to extend the results of the thesis to non-linear systems and non-linear MOR.

In the **second part** of this thesis, we investigated multirate methods for systems of differential-algebraic equations (DAEs) with different dynamical behaviour. Here, additional algebraic constraints have to be fulfilled in each integration step. Usually, this is realized by applying implicit time integration methods at least for the algebraic subsystems. We derived two multirate methods for semi-explicit DAEs of index-1 from implicit and stiffly accurate Runge-Kutta schemes. We showed consistency and convergence for both multirate methods and proved convergence order 1 and 2, respectively. In the same way, other multirate integration methods for semi-explicit DAEs of index-1 based on implicit and stiffly accurate Runge-Kutta schemes can be derived.
The analysis of the derived multirate DAE-integration schemes yields two major results:
1. For the Coupled-Slowest-First approach, the computation during the macro-step is performed on a uniform time grid. This leads to higher accuracy of the resulting multirate DAE-integration scheme compared to the other coupling approaches.
2. During the computation of the micro-steps, the values of the coupling-variables are achieved by linear interpolation. Hereby, the convergence order of the resulting multirate DAE-integration scheme is bounded by 2.
Numerical Simulations confirmed the derived convergence orders.
The interpretation of a Decoupled-Slowest-First multirate one-step method as a dynamic iteration scheme opens up new perspectives in the analysis of multirate time integration schemes. In this thesis, we stated and proved a convergence theorem for a general multirate one-step method using the Decoupled-Slowest-First approach. Here, the convergence of the multirate method depends on the properties of the considered system of DAEs, but the theorem is valid for all multirate one-step methods using the Decoupled-Slowest-First approach.

References to current work and outlook:
The usage of linear interpolation to evaluate the coupling variables during the computation of the micro-steps turned out to be the bottleneck in the derivation of highly accurate multirate time integration methods for DAEs. For differential variables, interpolation formulas of higher order can be used [Sch20]. To interpolate the algebraic variables with higher accuracy, a numerical differentiation is necessary which leads to an additional computational effort and additional error terms. What remains is to develop an efficient, robust and reliable strategy to evaluate the algebraic coupling variables to derive multirate methods with a high convergence order.

The interpretation of multirate one-step methods as dynamic iteration schemes has already been advanced and first results have been published [BG20]. For further coupling strategies, the splitting functions have been formulated and the corresponding convergence theorems were derived. This approach enables to analyse existing multirate methods from a new point of view and to develop new ideas to improve the approximation properties of multirate methods.

Outlook – Model Order Reduction and Multirate Time Integration for DAEs

In this thesis, we applied an MOR to the slow changing subsystem of a multiscale ODE and we derived multirate time integration methods for DAEs. For future research, it is now quite natural to consider a coupled system of DAEs with a high dimensional slow subsystem, apply an MOR to the slow changing, DAE subsystem, integrate the coupled system with reduced order, slow subsystem and investigate computation time and approximation properties of the multirate method. In many technical applications, the mathematical model leads to a multiscale system of coupled DAEs with high dimensional, slow subsystem, e.g. highly integrated electrical circuits or

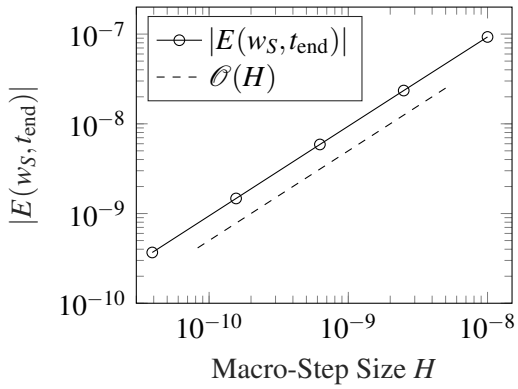
the field-circuit coupled system of Section 3.4. There exist MOR techniques for particular DAEs which usually exploit the special structure given by the underlying physics, e.g. [SS17] for circuit equations or [KBS17] for magneto-quasistatic Maxwell's equation. Applying an existing MOR technique to the slow changing DAE subsystem, a future field of research is the investigation of the impact of the MOR to the computation time and the approximation properties of a multirate integration schemes for DAEs.

A Appendix A

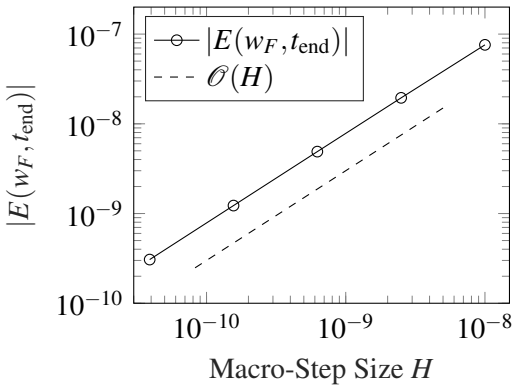
Convergence Plots of the mrIRK1-DAE scheme

In the following, the results of the numerical simulation for the mrIRK1-DAE scheme are given, c.f. Section 4.2.5. We show the convergence properties for each component of the DAE system (4.49) for all three coupling approaches: Decoupled-Slowest-First in Section A.1, Coupled-Slowest-First in Section A.2 and Coupled-First-Step in A.3.

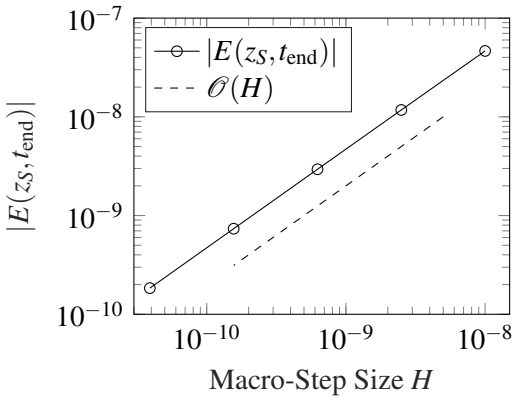
A.1 Decoupled-Slowest-First



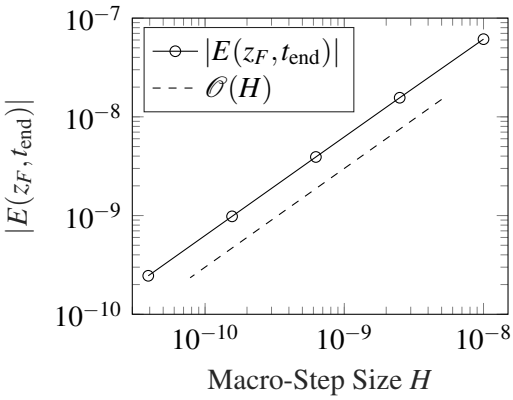
(a) $E(w_S,t_{\text{end}})$ in $\mathcal{O}(H)$



(b) $E(w_F,t_{\text{end}})$ in $\mathcal{O}(H)$



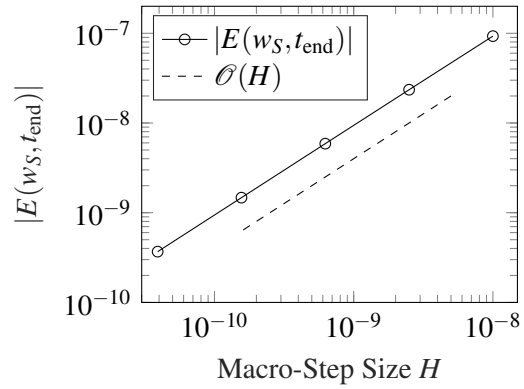
(c) $E(z_S,t_{\text{end}})$ in $\mathcal{O}(H)$



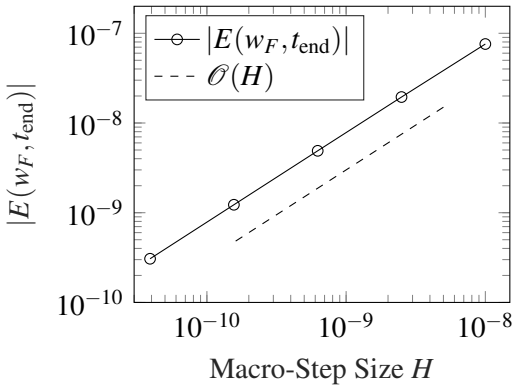
(d) $E(z_F,t_{\text{end}})$ in $\mathcal{O}(H)$

Figure A.1: Order of convergence of the mrIRK1-DAE scheme for the Decoupled-Slowest-First approach (m=10): a)-d) order 1.

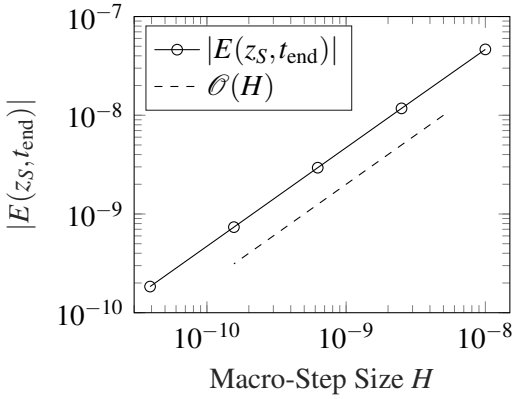
A.2 Coupled-Slowest-First



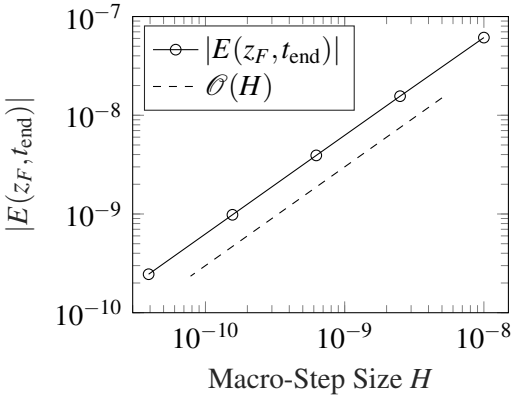
(a) $E(w_S,t_{\text{end}})$ in $\mathcal{O}(H)$



(b) $E(w_F,t_{\text{end}})$ in $\mathcal{O}(H)$



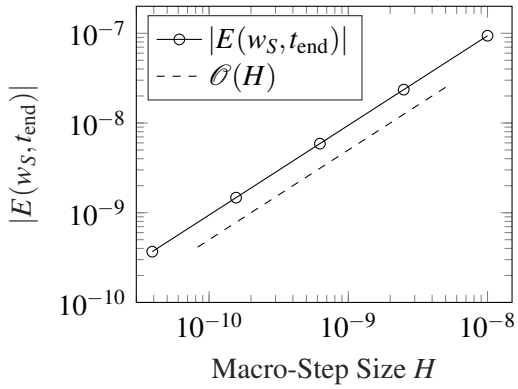
(c) $E(z_S,t_{\text{end}})$ in $\mathcal{O}(H)$



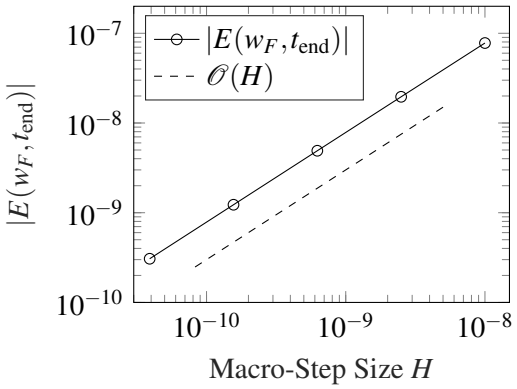
(d) $E(z_F,t_{\text{end}})$ in $\mathcal{O}(H)$

Figure A.2: Order of convergence of the mrIRK1-DAE scheme for the Coupled-Slowest-First approach (m=10): a)-d) order 1.

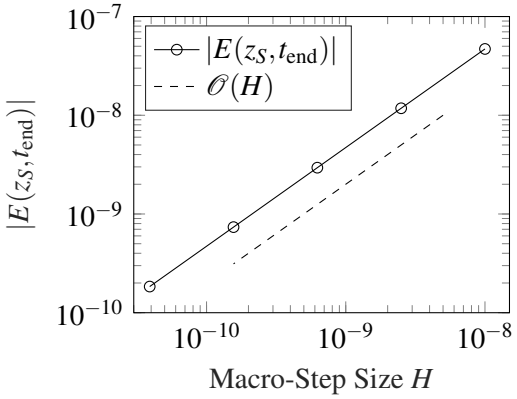
A.3 Coupled-First-Step



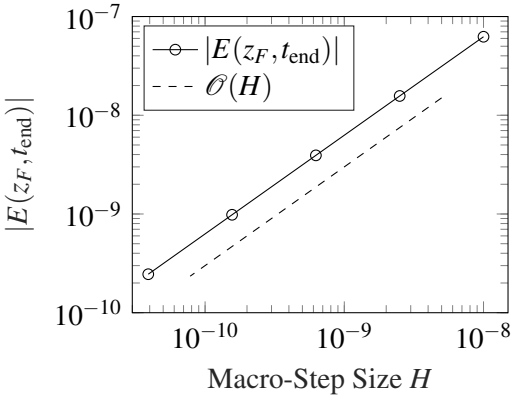
(a) $E(w_S,t_{\text{end}})$ in $\mathscr{O}(H)$



(b) $E(w_F,t_{\text{end}})$ in $\mathscr{O}(H)$



(c) $E(z_S,t_{\text{end}})$ in $\mathscr{O}(H)$



(d) $E(z_F,t_{\text{end}})$ in $\mathscr{O}(H)$

Figure A.3: Order of convergence of the mrIRK1-DAE scheme for the Coupled-First-Step approach (m=10): a)-d) order 1.

References

[AG01] M. Arnold and M. Günther, *Preconditioned dynamic iteration for coupled differential-algebraic systems*, BIT Numerical Mathematics **41** (2001), no. 1, 1–25.

[Ant05] A. C. Antoulas, *Approximation of large-scale dynamical systems (advances in design and control)*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.

[Bar01] A. Bartel, *Multirate ROW methods of mixed type for circuit simulation*, Scientific Computing in Electrical Engineering. Lecture Notes in Computational Science and Engineering (Berlin) (U. van Rienen and Michael Günther, eds.), Springer, 2001, pp. 241 – 249.

[Bar04] _____, *Partial differential-algebraic models in chip design - thermal and semiconductor problems*, PhD. thesis, Technische Universität München. Fortschritt-Berichte VDI, Reihe 20, VDI-Verlag Düsseldorf, 2004.

[BBGS13] A. Bartel, M. Brunk, M. Günther, and S. Schöps, *Dynamic iteration for coupled problems of electric circuits and distributed devices*, SIAM J. Sci. Comput. **35** (2013), no. 2, B315–B335.

[BBS14] A. Bartel, M. Brunk, and S. Schöps, *On the convergence rate of dynamic iteration for coupled problems with multiple subsytems*, Journal of Computational and Applied Mathematics **262** (2014), 14–24.

[BCG20] M.W.F.M. Bannenberg, A. Ciccazzo, and M. Günther, *Coupling of model order reduction and multirate techniques for coupled dynamical systems*, IMACM Preprint No. 22/2020, Bergische Universität Wuppertal, Submitted to Applied Mathematics Letters, 2020.

[BCP95] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical solution of initial-value problems in differential-algebraic equations*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1995.

[BG02] A. Bartel and M. Günther, *A multirate W-method for electrical networks in state-space formulation*, Journal of Computational and Applied Mathematics **147** (2002), no. 2, 411 – 425.

[BG20] _____, *Inter/extrapolation-based multirate schemes – a dynamic-iteration perspective*, 2020, arXiv:2001.02310.

[BGK02] A. Bartel, M. Günther, and A. Kværnø, *Multirate methods in electrical circuit simulation*, Progress in Industrial Mathematics at ECMI 2000. Mathematics in Industry (A. M. Anile, V. Capasso, and A. Greco, eds.), no. 1, Springer Berlin Heidelberg, 2002, pp. 258 – 265.

[BGS03] A. Bartel, M. Günther, and M. Schulz, *Modeling and discretization of a thermal-electric test circuit*, Modeling, Simulation and Optimization of Integrated Circuits. International Series of Numerical Mathematics (K. Antreich, R. Bulirsch, A. Gilg, and Peter Rentrop, eds.), 146, Birkhäuser, 2003, pp. 187 – 201.

[BS13] P. Benner and J. Saak, *Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state of the art survey*, GAMM-Mitteilungen **36** (2013), no. 1, 32–52.

[Bur95] K. Burrage, *Parallel and sequential methods for ordinary differential equations*, Clarendon Press, Oxford, 1995.

[But64] J. C. Butcher, *Implicit Runge-Kutta processes*, Mathematics of Computations **18** (1964), 50–64.

[CG95] S. L. Campbell and C. W. Gear, *The index of general nonlinear DAEs*, Numerische Mathematik **72** (1995), no. 2, 173–196.

[CS10] E. Constantinescu and A. Sandu, *On extrapolated multirate methods*, Progress in Industrial Mathematics at ECMI 2008 (Alistair D. Fitt, John Norbury, Hilary Ockendon, and Eddie Wilson, eds.), Mathematics in Industry, Springer Berlin Heidelberg, 2010, pp. 341–347 (English).

[CS12] S. Chaturantabut and D. C. Sorensen, *A state space error estimate for POD-DEIM nonlinear model reduction*, SIAM J. Numerical Analysis **50** (2012), no. 1, 46–63.

[Dah59] G. Dahlquist, *Stability and error bounds in the numerical integration of ordinary differential equations*, Handlingar, Almquist & Wiksells boktr. [H. Lindståhls bokhandel i distribution, Stockholm], 1959.

[DHZ87] P. Deuflhard, E. Hairer, and J. Zugck, *One-step and extrapolation methods for differential-algebraic systems*, Numerische Mathematik **51** (1987), no. 5, 501–516.

[EL97] Ch. Engstler and Ch. Lubich, *Multirate extrapolation methods for differential equations with different time scales.*, Computing **58** (1997), no. 2, 173–186.

[ESF98] E. Eich-Soellner and C. Führer, *Numerical methods in multibody dynamics*, European Consortium for Mathematics in Industry, Vieweg+Teubner Verlag, 1998.

[GF99] M. Günther and U. Feldmann, *CAD based electric circuit modeling in industry. Part I: Mathematical structure and index of network equations. Part II: Impact of circuit configurations and parameters*, Surv. on math. for ind. **8** (1999), 97–129 (english).

[GK01] M. Günther and A. Kværnø, *Multirate partitioned Runge-Kutta methods*, BIT **41** (2001), no. 3, 504 – 515.

[GM86] E. Griepentrog and R. März, *Differential-algebraic equations and their numerical treatment*, Teubner-Texte zur Mathematik, Teubner Verlag, 1986.

[GR93] M. Günther and P. Rentrop, *Multirate ROW methods and latency of electric circuits*, Applied Numerical Mathematics **13** (1993), no. 1-3, 83 – 102.

[Gru19] G. Gruebl, *Mathematische Methoden der Theoretischen Physik 1*, Springer Spektrum,

Berlin Heidelberg, 2019.

[GS16] M. Günther and A. Sandu, *Multirate generalized additive Runge-Kutta methods*, Numerische Mathematik **133** (2016), no. 3, 497–524.

[GS20] _____, Private Communication, 2018–2020, Content of the communication was a book project about multirate methods which will be published in the near future.

[GW84] C. W. Gear and D. R. Wells, *Multirate linear multistep methods*, BIT **24** (1984), no. 4, 484–502.

[HBG16a] C. Hachtel, A. Bartel, and M. Günther, *Interface reduction for multirate ODE-solver*, Scientific Computing in Electrical Engineering at SCEE 2014, Wuppertal, Germany, July 2014 (Andreas Bartel, Marcus Clemens, Michael Günther, and E. Jan W. ter Maten, eds.), Springer Berlin Heidelberg, 2016.

[HBG16b] _____, *Model order reduction for multirate ODE-solvers in a multiphysics application*, Progress in Industrial Mathematics at ECMI 2014 (Giovanni Russo, Vincenzo Capasso, Giuseppe Nicosia, and Vittorio Romano, eds.), Springer Berlin Heidelberg, 2016.

[HBG+18] C. Hachtel, A. Bartel, M. Günther, J. Kerler-Back, and T. Stykel, *Multirate DAE/ODE-simulation and model order reduction for coupled field-circuit systems*, Scientific Computing in Electrical Engineering at SCEE 2016, St. Wolfgang, Austria, October 2016 (Ulrich Langer, Wolfgang Amrhein, and Walter Zulehner, eds.), Springer Berlin Heidelberg, 2018.

[HBGS19] C. Hachtel, A. Bartel, M. Günther, and A. Sandu, *Multirate implicit Euler schemes for a class of differential–algebraic equations of index-1*, Journal of Computational and Applied Mathematics (2019), 112499.

[HNW08] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving ordinary differential equations I: Nonstiff problems*, Springer Berlin Heidelberg, 2008.

[HS09] W. Hundsdorfer and V. Savcenco, *Analysis of a multirate theta-method for stiff ODEs*, Applied Numerical Mathematics **59** (2009), 693 – 706.

[HW02] E. Hairer and G. Wanner, *Solving ordinary differential equations II: Stiff and differential-algebraic problems*, Springer Berlin Heidelberg, 2002.

[KBS17] J. Kerler-Back and T. Stykel, *Model reduction for linear and nonlinear magneto-quasistatic equations*, International Journal for Numerical Methods in Engineering **111** (2017), no. 13, 1274–1299.

[KM06] P. Kunkel and V.L. Mehrmann, *Differential-algebraic equations: analysis and numerical solution*, EMS textbooks in mathematics, European Mathematical Society Zürich, 2006.

[KR99] A. Kværnø and P. Rentrop, *Low order multirate Runge-Kutta methods in electric circuit simulation*, Preprint No. 2/99, NTNU Trondheim, 1999.

[Kvæ00] A. Kværnø, *Stability of multirate Runge-Kutta schemes*, International Journal of Differential Equations and Applications **1A** (2000), 97–105.

[Mär02] R. März, *The index of linear differential algebraic equations with properly stated leading terms*, Results in Mathematics **42** (2002), 308–338.

[Meh15] V. Mehrmann, *Index concepts for differential-algebraic equations*, Encyclopedia of Applied and Computational Mathematics (B. Engquist, ed.), Springer Berlin Heidelberg, 2015, pp. 676–681.

[Ple77] R. J. Plemmons, *M-matrix characterizations. I – nonsingular M-matrices*, Linear Algebra and its Applications **18** (1977), no. 2, 175–188.

[Ric60] J. R. Rice, *Split Runge-Kutta method for simultaneous equations*, Journal of Research of the National Bureau of Standards **64B** (1960), 151–170.

[RS07] T. Reis and T. Stykel, *Stability analysis and model order reduction of coupled systems*, Mathematical and Computer Modelling of Dynamical Systems **13** (2007), no. 5, 413–436.

[Sch11] S. Schöps, *Multiscale modeling and multirate time-integration of field/circuit coupled problems*, PhD. thesis, University of Wuppertal. Fortschritt-Berichte VDI, Reihe 21, Nr. 398, VDI-Verlag Düsseldorf, 2011.

[Sch20] K. Schäfers, *Analyse des Einsatzes inter-/extrapolierter Daten zur Konstruktion von Multirate Einschritt-Verfahren höherer Ordnung*, Bachelor thesis, Bergische Universität Wuppertal, 2020.

[SHV07] V. Savcenco, W. Hundsdorfer, and G. J. Verwer, *A multirate time stepping strategy for stiff ordinary differential equations*, BIT Numerical Mathematics **47** (2007), no. 1, 137–155.

[Sim16] V. Simoncini, *Computational methods for linear matrix equations*, SIAM Review **58** (2016), no. 3, 377–441.

[SS13] A. Steinbrecher and T. Stykel, *Model order reduction of nonlinear circuit equations*, International Journal of Circuit Theory and Applications **41** (2013), no. 12, 1226–1247.

[SS17] ———, *Element-based model reduction in circuit simulation*, System Reduction for Nanoscale IC Design (P. Benner, ed.), Springer, Berlin Heidelberg, 2017, pp. 39–85.

[Str06] M. Striebel, *Hierarchical mixed multirating for distributed integration of DAE network equations in chip design*, PhD. thesis, University of Wuppertal. Fortschritt-Berichte VDI, Reihe 20, Nr. 404, VDI-Verlag Düsseldorf, 2006.

[Ver08] A. Verhoeven, *Redundancy reduction of IC models by multirate time-integration and model order reduction*, Ph.D. thesis, Technische Universiteit Eindhoven, 2008.