A family of Newmark-type methods for singularly perturbed mechanical systems

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der
Naturwissenschaftlichen Fakultät II
der
Martin-Luther-Universität Halle-Wittenberg
Institut für Mathematik

vorgelegt von
Herrn Markus Arthur Köbis
geboren am 30. September 1985
in Halle (Saale)

Gutachter: Prof. Dr. Martin Arnold
Prof. Dr. Bernd Simeon

Betreuer: Prof. Dr. Martin Arnold
Tag der Verteidigung: 20. Oktober 2016

Halle (Saale), den 30. Juni 2016

Contents

Danksagung

Als erstes möchte ich an dieser Stelle dem Betreuer meines Promotionsvorhabens Prof. Dr. Martin Arnold danken. Er hat mir die Fragestellungen dieser Dissertation nahegebracht und war über viele Jahre eine große Stütze. Diese Arbeit wäre ohne seine kontinuierliche Unterstützung und seine fachlichen Ratschläge nicht möglich gewesen.

Auch den anderen aktuellen und ehemaligen Mitgliedern der „Arbeitsgruppe Numerik“ der Martin-Luther-Universität möchte ich hiermit für die stets angenehme Arbeitsatmosphäre und Hilfsbereitschaft danken. Ich werde immer mit Freude an meine Zeit am Institut zurückdenken.

Den Mitgliedern unseres „Arbeitsgruppenseminars Numerische Mathematik“ danke ich für die vielen anregenden Diskussionen und die Möglichkeit, Zwischenergebnisse meiner Arbeit regelmäßig vorstellen zu können und kritisch hinterfragt zu wissen. Insbesondere bei Dr. Helmut Podhaisky möchte ich mich für die Hilfestellungen und Ratschläge bedanken, die teilweise in diese Arbeit eingeflossen sind.

Meine Beschäftigung mit singulär gestörten Systemen in der Mehrkörperdynamik begann während meiner Mitarbeit im Verbundprojekt „SNiMoRed“ im Rahmen des Programms zur Förderung der Grundlagenforschung auf dem Gebiet „Mathematik für Innovationen in Industrie und Dienstleistungen“ des Bundesministeriums für Bildung und Forschung. Allen Beteiligten dieses Projekts möchte ich ebenso meinen Dank aussprechen. Zudem bedanke ich mich für die Zuwendungen durch die Stiftung Theoretische Physik/Mathematik, die mir mehrere Forschungsreisen im Rahmen meiner Promotion ermöglicht haben.

Vor allen Dingen aber möchte ich meiner Frau Elisabeth für ihre immerwährende uneingeschränkte Unterstützung danken.

Chapter 1

Motivation

1.1 A prominent example

We will start the investigation by considering the 'inevitable' (quote E. Hairer at the Numdiff-13 conference, see also (Simeon, 2015)) pendulum example: Consider a body whose mass m is condensed to a single point, moving in a plane under the influence of gravitational force $m \cdot g_{\text{grav}} \cdot (0,-1)^\top$ and attached to the origin of the coordinate system (q_x, q_y) by a mass-less rod of length l, cf. Figure 1.1. Although it is possible to describe the motion of the body in time t by means of the angle $\varphi_{\min}$ between rod and q_y-axis we will (e. g. for reasons of easier geometric interpretation) use the given coordinate system. Postponing the derivation to the next chapter we simply state that the motion can be described by the following 'mixture' of differential and algebraic equations.

$$\begin{aligned} m\ddot{q}_x(t) &= & &-\frac{q_x(t)}{\sqrt{(q_x(t))^2 + (q_y(t))^2}}\lambda(t)\,, \\ m\ddot{q}_y(t) &= -mg_{\text{grav}} & &-\frac{q_y(t)}{\sqrt{(q_x(t))^2 + (q_y(t))^2}}\lambda(t)\,, \end{aligned} \tag{1.1a}$$

$$g\left((q_x(t), q_y(t))\right) := \sqrt{(q_x(t))^2 + (q_y(t))^2} - l = 0\,. \tag{1.1b}$$

Here and throughout the entire work the superposed dot $\dot{(\bullet)} := \frac{\mathrm{d}}{\mathrm{d}t}(\bullet)$ denotes differentiation with respect to time. The first two equations in (1.1) reflect Newton's law that the product of acceleration of a body and its mass equals the force acting on that body. In the right-hand side of (1.1a) an additional variable $\lambda(t)$ is introduced which is necessary to assure that $g = 0$ can be attained for the solution. Approximating $\lambda(t)$ sufficiently well becomes an important task whenever one is interested in exactly calculating the forces in the system, i. e., the force on the rod. Analytically there should at first sight be no difference when instead of $g = 0$ the third equation is replaced by its time derivative

$$\frac{\mathrm{d}}{\mathrm{d}t} g(q_x(t), q_y(t)) = \frac{q_x(t)\dot{q}_x(t) + q_y(t)\dot{q}_y(t)}{\sqrt{(q_x(t))^2 + (q_y(t))^2}}\,, \tag{1.1c}$$

but we will see in a moment that Newmark integrators, numerical time integration algorithms which are very popular in structural and multibody dynamics perform different when applied to the system in its different formulations.

As the additional algebraic equation in (1.1) spoils the mathematical structure as a purely differential equation one might be interested in describing the motion without such additional

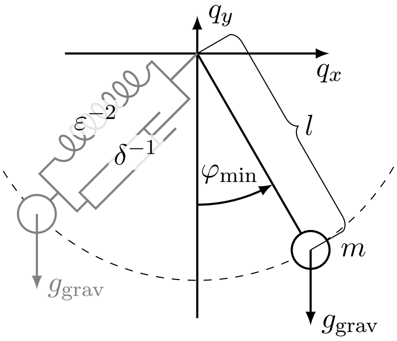


Figure 1.1: The mathematical pendulum

constraints. One approach (motivated by physics reasoning) might be to replace the rod by a (again mass-less) spring with very large stiffness, cf. Figure 1.1 (gray). If that spring constant is given by $\frac{1}{\varepsilon^2}$ for a (very small) constant $0 < \varepsilon \ll 1$ the equations may be stated as

$$\begin{aligned} m\ddot{q}_x^\varepsilon(t) &= & -\frac{q_x^\varepsilon(t)}{\varepsilon^2}\frac{\sqrt{(q_x^\varepsilon(t))^2+\left(q_y^\varepsilon(t)\right)^2}-l}{\sqrt{(q_x^\varepsilon(t))^2+\left(q_y^\varepsilon(t)\right)^2}}, \\ m\ddot{q}_y^\varepsilon(t) &= -mg_{\text{grav}} & -\frac{q_y^\varepsilon(t)}{\varepsilon^2}\frac{\sqrt{(q_x^\varepsilon(t))^2+\left(q_y^\varepsilon(t)\right)^2}-l}{\sqrt{(q_x^\varepsilon(t))^2+\left(q_y^\varepsilon(t)\right)^2}}. \end{aligned} \tag{1.2}$$

From physical intuition one might nevertheless (correctly) expect the solution of (1.2) to be vibrating, i. e., to show high frequency, yet low amplitude oscillations, which is rather challenging for numerical time integration schemes. So, in a fourth attempt the spring can be replaced by a damper element with damping coefficient $\frac{1}{\delta}$, $0 < \delta \ll 1$. Omitting the derivation once more, the motion in that case is described by the differential equation

$$\begin{aligned} m\ddot{q}_x^\delta(t) &= & -\frac{q_x^\delta(t)}{\delta}\cdot\frac{q_x^\delta(t)\dot{q}_x^\delta(t)+q_y^\delta(t)\dot{q}_y(t)}{(q_x^\delta(t))^2+(q_y^\delta(t))^2}, \\ m\ddot{q}_y^\delta(t) &= -mg_{\text{grav}} & -\frac{q_y^\delta(t)}{\delta}\cdot\frac{q_x^\delta(t)\dot{q}_x^\delta(t)+q_y^\delta(t)\dot{q}_y(t)}{(q_x^\delta(t))^2+(q_y^\delta(t))^2}. \end{aligned} \tag{1.3}$$

Let us assume that in its initial configuration the mass $m = 1$ is situated at $(q_x, q_y) = \frac{1}{2}(\sqrt{2}, \sqrt{2})$, and has a velocity $(\dot{q}_x, \dot{q}_y) = (-1, 1)$ (physical units omitted) for a given rod length $l = 1$ and gravitational acceleration $g_{\text{grav}} = 9.81$. We will apply the *generalized-α* (CH(0.7)) algorithm of Chung and Hulbert (1993), a member of the above mentioned Newmark family to the problem to get an approximation to the position, velocity and acceleration coordinates of the point mass. The constants of the physically inspired equations are set to $\varepsilon^2 = \delta = 10^{-6}$. Having a closer look at the substitute problems (spring or damper system, respectively) we observe that known quantities

$$(q_x^{\varepsilon|\delta}, q_y^{\varepsilon|\delta}, \dot{q}_x^{\varepsilon|\delta}, \dot{q}_y^{\varepsilon|\delta})$$

can be used to get approximations to the variables $\lambda = \lambda^{\varepsilon|\delta}$ using the relations

$$\lambda^{\varepsilon} := \frac{1}{\varepsilon^2}\left(\sqrt{(q_x^{\varepsilon})^2 + (q_y^{\varepsilon})^2} - l\right) ,$$
$$\lambda^{\delta} := \frac{1}{\delta}\frac{q_x^{\delta}\dot{q}_x^{\delta} + q_y^{\delta}\dot{q}_y}{\sqrt{(q_x^{\delta})^2 + (q_y^{\delta})^2}} ,$$

such that an estimate of the correct forces on the rod (in the fixed length model) can also be acquired if the substitute models are solved.

In Figure 1.2 the errors in the variable λ, i. e., the difference of a reference solution of (1.1) that has been obtained by using a standard time integration method with very low tolerances and the numerical approximations for step size $h = 2 \cdot 10^{-3}$ and the four different formulations are illustrated as they evolve throughout the time integration. The upper plot collects all the data, for a better understanding in the lower four plots we zoomed in and made the data points not in focus transparent to provide a better resolution of the transient phase.

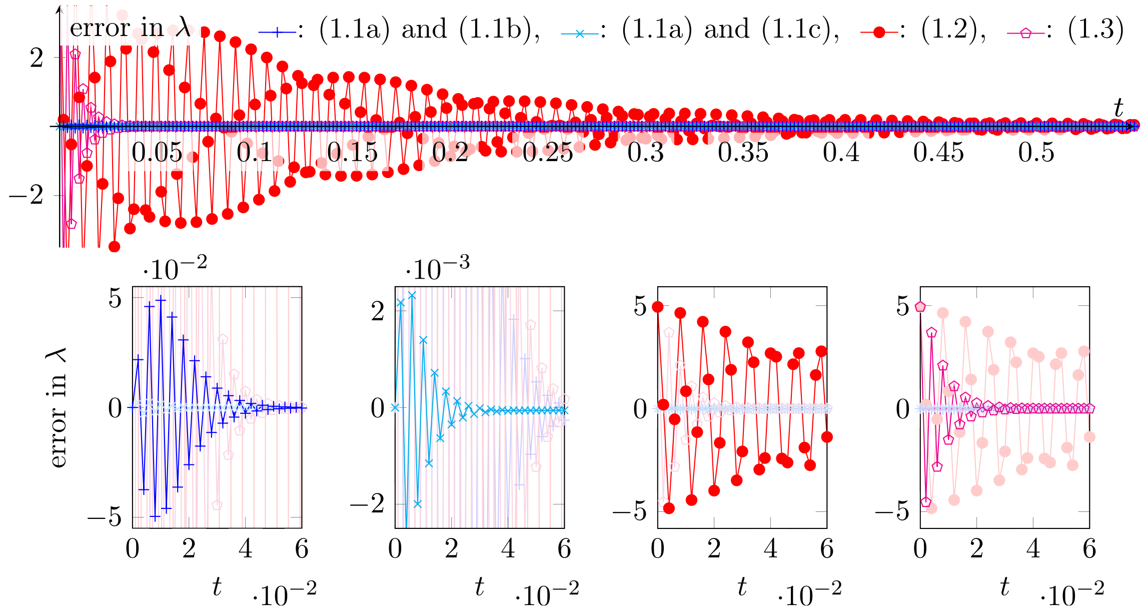


Figure 1.2: Errors in variable λ of the Chung–Hulbert(0.7) integrator for the pendulum benchmark with position or velocity constraint and the spring or damper substitute model

We hereby tacitly assume that obtaining the numerical solution did not cause any further difficulties. One peculiar—and probably the most important—feature of Newmark-type integrators can immediately be recognized: *Numerical damping* is the ability of the scheme to damp out error terms or unnatural, so-called *spurious oscillations* and hereby most importantly stabilize the time integration process itself. Straightaway, one can also see that the error terms in the two substitute formulations are substantially bigger in the first time steps and that for the original (index-3) and the spring model the errors are not damped out instantaneously but instead increase in the first time steps before the numerical damping effect sets in.

Unfortunately, for large scale simulations in industrial implementations we are not in the undoubtedly favorable position to simply compute initial values that fulfill the constraint equations exactly. It can, on the contrary, even be the case that the constraint equations are not even known. Figure 1.3 illustrates the results in this situation: Using again the constant time step size $h = 2 \cdot 10^{-3}$ we performed a series of experiments for the two substitute problems (s: spring, d: damper model) where the initial values have been perturbed like

$$(q_x^{\varepsilon|\delta}, q_y^{\varepsilon|\delta})(0) = \big((q_x, q_y)(0)\big) \cdot (1 + \Delta_q)\,, \quad (\dot{q}_x^{\varepsilon|\delta}, \dot{q}_y^{\varepsilon|\delta})(0) = (\dot{q}_x, \dot{q}_y)(0) + \Delta_v \cdot (1,1)\,.$$

The plot shows the error of the numerical results with respect to a reference solution of (1.1) dependent on the physical parameters ε and δ. As one would expect that error decreases as ε and δ become smaller; this is simply a smaller modeling error. Nonetheless, if the deviation from $g|_{t=0} = 0$ is in the magnitude of h^2 the spring model (1.2) does not approximate the original model sufficiently accurate anymore. This unfavorable behavior may be avoided if we change to the damper model (1.3). If the deviation from $\frac{\mathrm{d}}{\mathrm{d}t} g|_{t=0} = 0$ becomes larger, even the damper model does no longer converge.

The numerical experiments give rise to the following questions:

(a) Why and how is the influence of the initial values and initialization of the algorithm so important? Why are there ‘humps’ in the initial errors that are later on damped out anyway and are there ways to prevent this deficiency?

(b) Does the algorithm converge for general mechanical systems in comparable formulations and if so which conditions on the initial values are necessary? What is the order of convergence measured in both: time step size and parameter ε or δ, respectively?

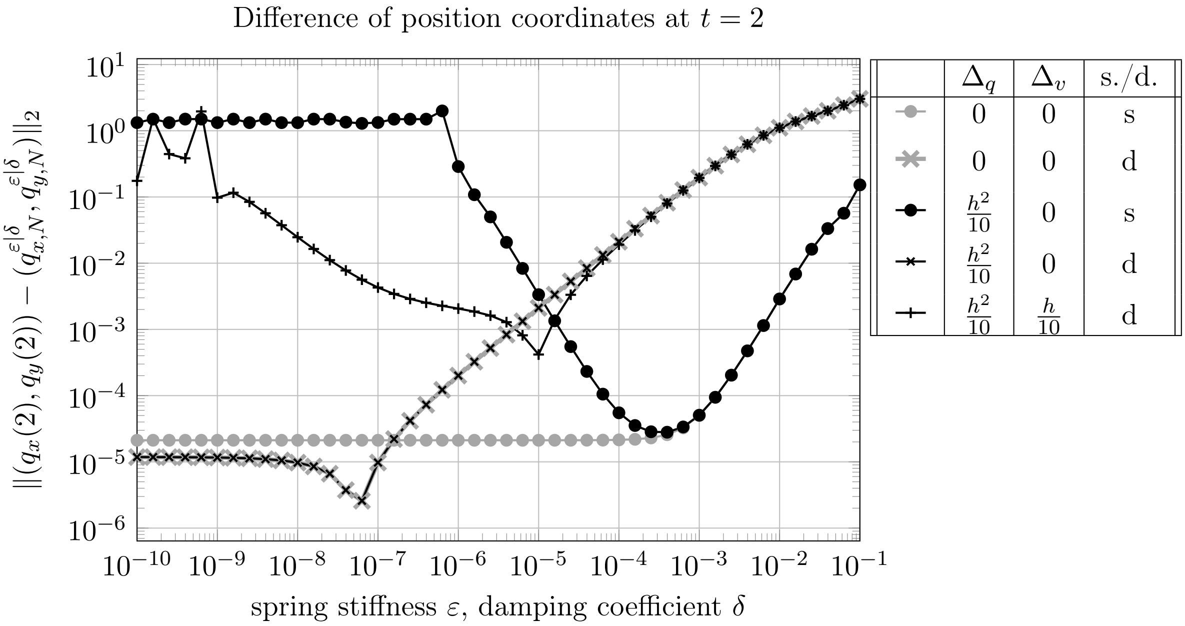


Figure 1.3: Deviation of position coordinates for the planar pendulum at final time $t = 2$ from reference solution (fixed length pendulum) for disturbed initial values, time step size $h = 2 \cdot 10^{-3}$.

This thesis is devoted to the application and analysis of *Newmark-type* (or generalized-α) time integration methods in the context of mechanical systems that, as the very simple pendulum example, are subject to constraints on either position or velocity coordinates which are enforced by including these constraints ‘as they are’ or by large additional force terms in the model, so-called *singularly perturbed systems.* Starting with elementary properties of the method for linear systems we will present a very broad convergence analysis which enables us to fully understand the above observations.

1.2 Outline

To make this thesis self-content we start by introducing the very basic concepts from mechanics and time integration methods in Chapter 2. Basic physical principles are used to derive a general setting on which the mechanical models are based. Systems with position constraints are characterized as *index*-3, a certain type of velocity-constrained systems as index-2 *differential-algebraic equations.*

Using the same principles we introduce a general framework for the two substitute problems in Chapter 3 and classify them as singularly perturbed problems. Important analytic properties resulting in the fundamental Rubin–Ungar Theorem are collected to gain a deeper understanding of the mathematical structure. The theoretical findings are underlined by nontrivial analytic and numerical examples.

Chapter 4 shall serve as an attempt to give a comprehensive overview on the developments of Newmark-type methods in their algorithmical and application context. Here, we establish the (mostly linear) theory of Newmark-type time integration methods and embed them within other methods in technical simulation. The main focus lies on stability and the closely related feature of controllable damping but we will also be able to understand the cause of spurious oscillations and derive a new parameter set to decrease these artifacts.

In Chapter 5 convergence of the Newmark integrators in the constrained case (index-3 and 2) as well as the singularly perturbed setting is demonstrated. Hereby, we emphasize the close relations but also distinctions of constrained vs. singularly perturbed problems on the one hand and index-2 vs. 3 and damper vs. spring model on the other hand.

Chapter 6 is a collection of a series of important issues and solutions concerning a practical implementation of the methods where especially the efficient solution of corrector equations in each step is addressed. Some numerical tests for benchmarks of small and moderate size are presented that verify the convergence theory before we summarize our findings in Chapter 7 and give a short outlook on topics of possible further research.

Chapter 2

Preliminaries

In the design and validation process of technical systems throughout all branches of science and engineering the need for reliable simulations prior to the real-world construction is increasingly important. Not only due to the high costs of physical prototyping and testing but also because of the always new emerging technical improvements and market demands in our accelerating world it is important to rely on numerical models and their efficient computational solution.

A particular challenge for the technical simulations lies in the robust treatment of (large) systems of ordinary differential equations (ODEs) which can be characterized by very different time scales. Especially in the field of biomechanical simulation, which has steadily been growing throughout the last decade, two problem classes of this type are of particular interest:

On the one hand, *stiff mechanical systems* always appear when large potential forces push the system in such a way that certain configurations become prohibitively improbable since they can only be reached utilizing a large amount of energy. The solutions of stiff mechanical systems show a typical behavior of large-frequencies and low-amplitude oscillations (vibrations) which are particularly challenging for most numerical solution procedures regardless of the fact that a good resolution of the vibrations of the system is actually not necessary at all. On the other hand, *strongly damped mechanical systems* are characterized by their ability to push the systems away from those high energy states (including the immense energy loss). Their typical solution behavior is therefore characterized by a short transient phase with large forces in the system which presents a challenge for many time integration algorithms as well.

In view of the application field of biomechanics, both categories are of cardinal importance as the modeling of biological tissue is often based on equations of structural analysis (Hughes, 1987, Simeon et al., 2009) for almost incompressible materials. Models of human joints or wobbling masses for interaction models need to take into account that living bodies are to some extend designed to be shock absorbent and energy dissipating for otherwise the risk of injury for regular interaction with the natural environment would be too high (Hans, 2004).

Either one of the two problem classes falls into the field of *singularly perturbed problems* (SPPs) and so they are not encompassed by classical convergence theory for the numerical solution of ODEs. In fact, it is well-known (Hairer and Wanner, 2002) that the study of numerical schemes for SPPs is almost unavoidably intertwined with the theory of differential-algebraic equations (DAEs). So, in this introductory chapter we will give an overview on basic properties and concepts of DAEs in mechanical system simulation and their numerical time integration, starting with the basic physical laws determining the time evolution of most mechanical systems.

This chapter is to a great extent inspired by the review article by Arnold et al. (2011) and the monograph by Eich-Soellner and Führer (1998).

2.1 Multibody system models and differential-algebraic equations

Of course, the mathematical description of structures in the physical world by a set of mathematical equations is always an abstraction from the real world. Since setup of industrially and commercially appropriate mock-ups are no topic to be answered by mathematical analysis and to circumscribe the scope of the present work, we start by formally defining what the term ‘multibody system’ here shall stand for:

Definition 2.1 (Multibody system (Eich-Soellner and Führer, 1998))**.**
A multibody system (model) is characterized as the collection of a finite number of (first off rigid) bodies that are interconnected to each other by mass-less joints, bearings and force elements, such as springs, dampers or actuators.

The first step in the modeling process of a mechanical multibody system is the definition of a set of *generalized coordinates*

$$\boldsymbol{q} = \boldsymbol{q}(t) \in \mathbb{R}^{n_{\boldsymbol{q}}}, \;\; t \in [t_0, t_{\text{end}}] \,,$$

i. e., a set of continuous values that precisely define the state of the system at any point in time t. For the pendulum example from the first chapter, we have already seen that the number of coordinates is not fixed but depends on the engineer’s accuracy requirements, experience and intuition. We could for example also consider a flexible pendulum with bending of the rod.

Remark 2.2 (Flexible multibody systems)
Due to an emerging demand for lightweight models (e. g. in spacecraft engineering) and high precision mechanisms (e. g. in optomechanics and/or medical applications) the simplification to just rigid bodies is often too restrictive and more sophisticated models from structural, thermo, electro or fluid dynamics need to be taken into account as well (so-called multiphysics problems). From the mathematical viewpoint this leads to coupled systems of DAEs and partial differential equations (PDEs). For the discourse in this thesis, we will always assume that $\boldsymbol{q}(t) \in \mathbb{R}^{n_{\boldsymbol{q}}}$ is finite-dimensional. This does not exclude PDE models from the above problem complexes, on the contrary, but disregards the space-discretization of those models, such that we assume that these models are already in the form of a (large) system of ODEs. We will shortly exemplify the semidiscretization of PDEs using the Navier–Lamé equations of structural analysis in Example 3.19 below.

2.1.1 The equations of motion of dynamical systems in technical simulation

In simulation environments for multibody systems (see Schiehlen, 1990) the equations of motion are usually not obtained as such but methods from graph theory are used to define the topology of the system using the bodies as nodes/vertices connected by joints and force elements represented by the edges of the graph. This has the advantage that for tree-structured systems (i. e., those where the corresponding graph includes no simple cycles) it is possible to get the dynamic equations with a computational effort that grows only linearly with the number of bodies in the system (so-called O(N)-formalisms, see for instance (Lubich et al., 1992)). From the perspective of numerical mathematics, those O(N)-formalisms may be regarded as a (local, body-oriented) block Gauss elimination of the implicit equations of motion to be defined below. The mathematical foundation of this topologic approach is the theory of bond-graphs: Seen as a ’cause and effect’-system, the connection of two bodies implements a causality assignment. The unifying modeling language Modelica has been designed to take this matter into account when describing the mathematical structure of systems not only from technical simulation but also in a more general sense (Elmqvist et al., 1998, Olsson et al., 2012). The topological approach saves a lot of computer memory (which also grows (just) linearly with the number of bodies). The underlying physical principles are nevertheless constitution respectively conservation laws using local (between pairwise two connected nodes in the graph) equilibria of forces and torques in the mechanical system. Industrial codes sometimes even neglect certain terms, e. g. very small accelerations, to improve the efficiency and memory demands of their models accepting the actual violation of the physical laws. For the engineer, the design principles (and therefore the equations as well) are hidden anyway and the programs mostly can be reckoned as 'black boxes'.

For a brief overview on the physical principles, we will follow Arnold (1988) and Hairer and Wanner (2002) and use the *formalism of Lagrange* which is not at the basic of most multibody system simulation environments but allows for a compact description and a great generality. Suppose, that the *kinetic energy* $\mathcal{T}$ which comprises the contributions of (angular) velocity terms and the *potential energy* $\mathcal{V}$ consisting of terms that stem from the configuration of the system itself, e. g. the position in a gravitational field or bending energy in a flexible body, are given. The *Lagrange function* or *Lagrangian* $\mathcal{L}$ of the system is then defined by

$$\mathcal{L} := \mathcal{T} - \mathcal{V} \ .$$

If one assumes *conservative* (or *natural*) systems, i. e., energy conservation throughout time evolution, *Hamilton's principle* states that the time integral over the Lagrangian (the action integral) takes a stationary (i. e., almost inevitably minimal) value:

$$\int_{t_0}^{t_{\text{end}}} \mathcal{L}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)) \ \mathrm{d}t \to \text{stat.}$$

The *Euler equations* of variational analysis (or Lagrange equations of the second kind) provide a necessary condition for stationarity

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = \mathbf{0} \tag{2.1}$$

leading to the *equations of motion*

$$\mathbf{M}(\boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) = \boldsymbol{f}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t))\,, \tag{2.2}$$

where $\mathbf{M}(\boldsymbol{q}(t)) := \frac{\partial^2 \mathcal{T}}{\partial \dot{\boldsymbol{q}}^2}$ is the *mass* matrix (or inertia matrix) and $\boldsymbol{f}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)) := -\frac{\partial^2 \mathcal{T}}{\partial \dot{\boldsymbol{q}} \partial \boldsymbol{q}}\dot{\boldsymbol{q}} + \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}}$ the *generalized force vector.* In this formulation the underlying physical law—*Newton's law* that the product of mass and accelerations equals the sum of forces acting on a body—is apparent. For most systems in technical simulation there is no conservation of energy such that the right-hand side in (2.1) has to be modified by the addition of certain force terms, see Example 2.6 below. Because of its semi-explicit structure, (2.2) is much more practicable than (2.1) for designing numerical methods. Throughout this thesis we assume that the kinetic energy (locally) is a positive form in the velocity coordinates $\dot{\boldsymbol{q}}$. As a consequence, the mass matrix $\mathbf{M} \colon \mathbb{R}^{n_{\boldsymbol{q}}} \to \mathbb{R}^{n_{\boldsymbol{q}} \times n_{\boldsymbol{q}}}$ is always symmetric positive definite (at least in a neighborhood of the solution).

Remark 2.3 (Autonomous systems)
In the theoretical investigation of this work we will (mainly for the sake of brevity) restrict the analysis to the case of autonomous systems, i. e., those with mass $\mathbf{M}$, *force* $\boldsymbol{f}$ *(and constraint function* $\boldsymbol{g}$, *see below) being not explicitly dependent on the time variable* t.

Remark 2.4 (Hamiltonian systems)
The particular special case of mechanical systems we introduced so far is characterized by its energy conservation. This property of a mechanical multibody system bears many properties that may be exploited to design numerical schemes. Mathematically, for those systems it is possible to define a Hamilton function *or* Hamiltonian $\mathcal{H}$ *which (in classical mechanics) coincides with the total energy of the system*

$$\mathcal{H} := \mathcal{T} + \mathcal{V} \ .$$

The above equations of motion may then be stated as

$$\dot{\boldsymbol{p}}(t) = -\frac{\partial \mathcal{H}(\boldsymbol{q},\boldsymbol{p})}{\partial \boldsymbol{q}} \, , \quad \dot{\boldsymbol{q}}(t) = \frac{\partial \mathcal{H}(\boldsymbol{q},\boldsymbol{p})}{\partial \boldsymbol{p}} \, ,$$

where the generalized momenta variables $\boldsymbol{p} \in \mathbb{R}^{n_{\boldsymbol{q}}}$ have been introduced using the Legendre transformation (Nolting, 2006) and coincide in most cases with the actual (physical) momenta of the bodies. Throughout the past decade there has been a steady growth in research interest on Hamiltonian systems since analytic properties as conservation of energy/momenta or symplecticity allow for using highly sophisticated techniques. The first monograph to give an integral overview on methods and challenges for numerical methods applied to Hamiltonian problems is the one by Sanz-Serna and Calvo (1994). The Newmark integrators we investigate here are explicitly not designed for energy conservation but instead usually decrease energy (mainly that of vibrating substructures of the mechanical system); this numerical damping *will be studied in more detail in Chapter 4.*

Example 2.5 (Spring pendulum)
For the (mathematical) spring pendulum example from Chapter 1 the kinetic energy is given by $\mathcal{T} = \frac{m}{2}((\dot{q}_x(t))^2 + (\dot{q}_y(t))^2)$. The potential energy consists of the elevation energy of the pointmass $\mathcal{V}_1 = m g_{\text{grav}} q_y(t)$ and the energy necessary to stretch or compress the spring. Given the spring constant ε^{-2}, Hooke's law states $\mathcal{V}_2 = \frac{1}{2\varepsilon^2}\left(\sqrt{(q_x(t))^2 + (q_y(t))^2} - l\right)^2$, such that the Lagrangian is given by $\mathcal{L} = \mathcal{T} - \mathcal{V}_1 - \mathcal{V}_2$ and from the formalism one deduces with

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial(\dot{q}_x,\dot{q}_y)} &= (m\ddot{q}_x, m\ddot{q}_y)\, , \\ \frac{\partial \mathcal{L}}{\partial(q_x,q_y)} &= \left(-\frac{q_x}{\varepsilon^2}\frac{\sqrt{q_x^2+q_y^2}-l}{\sqrt{q_x^2+q_y^2}},\ -m g_{\text{grav}} - \frac{q_x}{\varepsilon^2}\frac{\sqrt{q_x^2+q_y^2}-l}{\sqrt{q_x^2+q_y^2}} \right)\, , \end{aligned}$$

that the equations of motion for the system are indeed given by (1.2). ♢

The decay of mechanical energy, or its transformation to e. g. thermal energy, is a very complex process. A unified treatment, like when deriving the equations of motion for Hamiltonian systems, is not as easily accomplished. To incorporate these effects in the mathematical models certain heuristics and linear surrogate models are mostly applied such that involved coefficients may be adapted to measurements. A very commonly used model will be explained in the following example.

Example 2.6 (Rayleigh-dissipation function (Nolting, 2006))
A prominent type of non-conservative forces in mechanical systems is realized by additional force terms that linearly depend on the generalized velocities. From a physical viewpoint they correspond to the concept of a *dissipation function of Rayleigh type*: Define an additional potential

term $\mathcal{D} := \frac{1}{2}\dot{\boldsymbol{q}}^\top \mathbf{D}\dot{\boldsymbol{q}}$ with a given symmetric positive semi-definite matrix $\mathbf{D} \in \mathbb{R}^{n_{\boldsymbol{q}} \times n_{\boldsymbol{q}}}$ and use the *modified Euler equations*

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = -\frac{\partial \mathcal{D}}{\partial \dot{\boldsymbol{q}}} \, .$$

It is easy to show that with this construction it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{T} + \mathcal{V}) = -\mathcal{D} \, ,$$

whenever the kinetic energy is a quadratic form in the generalized velocities $\dot{\boldsymbol{q}}$. So, the dissipation function may be seen as a measure for the energy loss (or its conservation) in the system. Note, that the dissipative forces are not present in the model if the generalized velocities $\dot{\boldsymbol{q}}$ lie in the nullspace of $\mathbf{D}$. In the literature, Rayleigh damping is occasionally simply referred to as ‘damped extension’ (of an undamped system). Sometimes, researchers only use the term ‘Rayleigh damping’ if $\mathbf{D} := d_1 \partial \boldsymbol{f}/\partial \boldsymbol{q} + d_2 \mathbf{M}$, d_1, $d_2 > 0$, is a linear combination of the mass matrix and the (tangent) stiffness matrix of the system (Hughes, 1987). This approach bears the advantage that only two parameters need to be fitted to the model and that for sufficiently smooth $\mathcal{L}$ a (local) diagonalization of the system is possible which is important for reduced order modeling (e. g. Craig–Bampton).

It is also possible to generalize the approach to variable (but still symmetric and positive semi-definite) matrices $\mathbf{D} = \mathbf{D}(t, \boldsymbol{q})$: If we consider once again the mathematical pendulum (see Chapter 1) the concept of a Rayleigh dissipation function can be used to obtain the dynamic equations in case of the (mass-less) damper system. To this end, we define

$$\mathbf{D}(q_x(t), q_y(t)) := \frac{1}{\delta} \cdot \frac{1}{(q_x(t))^2 + (q_y(t))^2} \begin{pmatrix} (q_x(t))^2 & q_x(t)q_y(t) \\ q_x(t)q_y(t) & (q_y(t))^2 \end{pmatrix} ,$$

such that $\mathbf{D}\dot{\boldsymbol{q}}$ vanishes if and only if the velocity vector $\dot{\boldsymbol{q}}$ and the pendulum’s rod are perpendicular. With that definition, we get exactly the equations (1.3) stated in the first chapter.

♢

2.1.2 Differential-algebraic equations

In case of mechanical multibody systems without tree structure it is often necessary to add algebraic conditions, *constraints*, that realize the closed loops in the topology while still allowing for global parameterization of the kinematics. The equations of motion can then be described as DAEs which in a very general form may be stated as

$$\boldsymbol{F}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)) = \mathbf{0}, \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \in \mathbb{R}^{n_{\boldsymbol{x}}} \, , \quad t \in [t_0, t_{\mathrm{end}}] \tag{2.3}$$

with a function $\boldsymbol{F} \colon \mathbb{R}^{2n_{\boldsymbol{x}}} \to \mathbb{R}^{n_{\boldsymbol{x}}}$ and a singular Jacobian $\mathbf{E} := \partial \boldsymbol{F}/\partial \dot{\boldsymbol{x}}$ with $0 < \operatorname{rank} \mathbf{E} < n_{\boldsymbol{x}}$. Including the two bounds just for a moment, we can already see that the analytic properties of solutions to DAEs range from that of ODEs ($\operatorname{rank} \mathbf{E} = n_{\boldsymbol{x}}$) to the solutions of nonlinear equations ($\operatorname{rank} \mathbf{E} = 0$). To classify the ‘difficulty’ of a DAE or how far away it is from the simple solution of an ODE, the concept of the *index* has been established.

Definition 2.7 (Differentiation index (Brenan et al., 1996))**.**
The minimum number of times that (possibly just parts of) (2.3) must be differentiated with respect to the time variable t in order to determine $\dot{\boldsymbol{x}}(t)$ as a continuous function of $\boldsymbol{x}$ (and t) is called the *differentiation index* of the DAE (if such a number exists).

The explicit form

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\chi}(t, \boldsymbol{x}(t)) \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{2.4}$$

determined from this differentiation procedure (*index-reduction*) is called the *underlying differential equation*. The somewhat vague formulation 'in order to determine' is in the literature sometimes also stated as 'such that algebraic manipulations allow for an extraction of an explicit form (2.4)' (cf. Hairer and Wanner, 2002). The assembly of $\boldsymbol{F}(\boldsymbol{x}, \dot{\boldsymbol{x}})$ and (all) its necessary time derivatives is called the *derivative array* of (2.3). From Definition 2.7 we see that for DAEs of differentiation index higher than one the time derivatives may imply additional algebraic conditions. These so-called *hidden constraints* also impose conditions on the initial values $\boldsymbol{x}_0$; DAEs of differentiation index higher than one are therefore considered as 'higher index DAEs'. Initial values fulfilling all (i. e., also hidden) constraints are said to be *consistent*. A direct application (or straightforward extension) of ODE time integration methods to higher index DAEs is inclined to fail or at least severely suffer from problems such as order reduction (Petzold, 1982, Brenan et al., 1996), see below. From that perspective, systems of differentiation index three fairly represent the frontier between well- and ill-posed problems and some researchers strongly recommend to perform analytic transformations or stabilization techniques instead of tackling these problems directly. With respect to that, the equations of motion of multibody system in descriptor form (see (2.12) below) are extraordinary: We will show that they are of differentiation index three but due to their special structure one can indeed (yet it is not common practice) treat them directly.

As it is possible (and most often the case anyway) to separate solely algebraic equations from the remaining part of a DAE, the following *semi-explicit* system can be seen as a rather natural special case:

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{\varphi}(\boldsymbol{y}(t), \boldsymbol{z}(t))\,, \tag{2.5a}$$

$$\mathbf{0} = \boldsymbol{\psi}(\boldsymbol{y}(t), \boldsymbol{z}(t))\,, \tag{2.5b}$$

where $\boldsymbol{x} =: (\boldsymbol{y}, \boldsymbol{z})^\top$ is a partitioning of $\boldsymbol{x}$ into *differential variables* $\boldsymbol{y}$ and *algebraic variables* $\boldsymbol{z}$.

Remark 2.8 (Index concepts in the literature)
We introduced the concept of the differentiation index in Definition 2.7 because it is the most commonly used and the easiest verifiable one. There are, nevertheless, some other ways to define the index of a DAE, and we enumerate the most important ones here. The starting point of DAE theory was probably the work of Gantmacher (1959) who studied linear, constant coefficient systems

$$\mathbf{E}_1\dot{\boldsymbol{x}}(t) + \mathbf{E}_2\boldsymbol{x}(t) = \boldsymbol{b}(t)\,. \tag{2.6}$$

A transformation of the matrix pencil $(\mathbf{E}_1, \mathbf{E}_2)$ *to* Weierstrass canonical form *allows to separate* $\mathbf{E}_1$ *into a regular and a nilpotent part. The order of nilpotency of the latter one defines the* nilpotency index. *Kunkel and Mehrmann (2006) introduced the* strangeness index *which allows for a direct extension to linear systems with time dependent coefficient matrices and also relies on the construction of an appropriate canonical form. From a functional analytic viewpoint, the most important index concept is the* perturbation index *(Hairer et al., 1989a). If one adds a (sufficiently small) function to the right-hand side(s) of* (2.3) *or* (2.5) *this index indicates the lowest order of derivatives of these perturbations which are needed to obtain an upper bound for the difference of the solutions to the original and the perturbed problem. Pantelides (1988) proposed an algorithm to detect loops in the dependency graphs of the variables and used that to define a* structural index *which is an appealing approach for multibody systems since* $O(N)$*-formalisms track these dependencies anyway.*

All these different index concepts have in common that (i) for linear systems (2.6) *they coincide (or at least can be treated equivalently) and (ii) they serve as a measure of how well the problem is posed and/or a measure of the distance to the ODE case. Whenever we just use the term 'index' we refer to the differentiation index of the system.*

Example 2.9 (Hessenberg systems (Clark (1988), Hairer and Wanner (2002, Sect. VII.1)))
If we consider the splitting given by (2.5) we can state explicit conditions such that the system is of index one or two: From a time differentiation of (2.5b) we obtain

$$\mathbf{0} = \boldsymbol{\psi}_{\boldsymbol{y}} \cdot \dot{\boldsymbol{y}} + \boldsymbol{\psi}_{\boldsymbol{z}} \cdot \dot{\boldsymbol{z}} = \boldsymbol{\psi}_{\boldsymbol{y}} \cdot \boldsymbol{\varphi}(\boldsymbol{y}, \boldsymbol{z}) + \boldsymbol{\psi}_{\boldsymbol{z}} \cdot \dot{\boldsymbol{z}}\,, \tag{2.7}$$

where $(\cdot)_{\boldsymbol{z}|\boldsymbol{y}}$ denotes partial derivatives and the time dependence is omitted for readability. If now

$$\boldsymbol{\psi}_{\boldsymbol{z}} \text{ is invertible,} \tag{2.8}$$

the Implicit Function Theorem allows for an explicit form to express $\dot{\boldsymbol{z}}$ in terms of $\boldsymbol{y}$ and $\boldsymbol{z}$. More precisely, in order to show also the existence of a solution on a finite time interval it is adequate even to postulate that the inverse $\boldsymbol{\psi}_{\boldsymbol{z}}$ remains bounded in a sufficiently large neighborhood of the (consistent) initial values. If (2.8) does not hold and $\boldsymbol{\psi}$ does not explicitly depend on $\boldsymbol{z}$, (2.7) reduces to the hidden constraint

$$\mathbf{0} = \boldsymbol{\psi}_{\boldsymbol{y}} \cdot \boldsymbol{\varphi}(\boldsymbol{y}, \boldsymbol{z})\,.$$

In this case

$$\text{regularity of } \boldsymbol{\psi}_{\boldsymbol{y}}\boldsymbol{\varphi}_{\boldsymbol{z}} \tag{2.9}$$

is a sufficient condition for index two since another differentiation then leads to

$$\big(\boldsymbol{\psi}_{\boldsymbol{y}}\boldsymbol{\varphi}_{\boldsymbol{z}}\big)\,\dot{\boldsymbol{z}} = -\boldsymbol{\psi}_{\boldsymbol{y}\boldsymbol{y}}\,(\boldsymbol{\varphi}, \boldsymbol{\varphi})\ , \quad \big((\partial(\boldsymbol{\psi}_{\boldsymbol{y}} \cdot \boldsymbol{w}_1)/\partial \boldsymbol{y})\boldsymbol{w}_2 =: \boldsymbol{\psi}_{\boldsymbol{y}\boldsymbol{y}}(\boldsymbol{w}_1, \boldsymbol{w}_2) \text{ for any } \boldsymbol{w}_1,\, \boldsymbol{w}_2 \in \mathbb{R}^{n_{\boldsymbol{y}}}\big)$$

such that under the stated regularity assumption (2.9) the system can be solved for $\dot{\boldsymbol{z}}$.

To further generalize to index-3 problems one is demanded to consider a different structure, namely Hessenberg systems of size three, i. e.,

$$\begin{aligned}
\dot{\boldsymbol{y}}^{(1)}(t) &= \boldsymbol{\varphi}^{(1)}(\boldsymbol{z}(t), \boldsymbol{y}^{(1)}(t), \boldsymbol{y}^{(2)}(t))\,,\\
\dot{\boldsymbol{y}}^{(2)}(t) &= \boldsymbol{\varphi}^{(2)}(\qquad \boldsymbol{y}^{(1)}(t), \boldsymbol{y}^{(2)}(t))\,,\\
\mathbf{0} &= \boldsymbol{\psi}\quad(\qquad\qquad\quad \boldsymbol{y}^{(2)}(t))\,.
\end{aligned}$$

Here, requiring that

$$\boldsymbol{\psi}_{\boldsymbol{y}^{(2)}} \cdot \boldsymbol{\varphi}^{(2)}_{\boldsymbol{y}^{(1)}} \cdot \boldsymbol{\varphi}^{(1)}_{\boldsymbol{z}} \text{ is invertible} \tag{2.10}$$

is a sufficient condition to obtain a system of index three, as can be seen by differentiating the last equation twice to get

$$\mathbf{0} = \boldsymbol{\psi}_{\boldsymbol{y}^{(2)}\boldsymbol{y}^{(2)}}(\boldsymbol{\varphi}^{(2)}, \boldsymbol{\varphi}^{(2)}) + \boldsymbol{\psi}_{\boldsymbol{y}^{(2)}}\boldsymbol{\varphi}^{(2)}_{\boldsymbol{y}^{(1)}}\boldsymbol{\varphi}^{(1)}\,,$$

which with $\bar{\boldsymbol{z}} := \boldsymbol{z}$, $\bar{\boldsymbol{y}} := (\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})^\top$, $\bar{\boldsymbol{\varphi}} := (\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})^\top$, $\bar{\boldsymbol{\psi}} := \boldsymbol{\psi}$ is the situation of the index-1 Hessenberg systems and $\bar{\boldsymbol{\psi}}_{\bar{\boldsymbol{z}}} = \boldsymbol{\psi}_{\boldsymbol{y}^{(2)}}\boldsymbol{\varphi}^{(2)}_{\boldsymbol{y}^{(1)}}\boldsymbol{\varphi}^{(1)}_{\boldsymbol{z}}$ is invertible by (2.10). ♢

Coming back to the simulation of mechanical systems we may require that the generalized coordinates $\boldsymbol{q}(t)$ fulfill a given set of $n_{\boldsymbol{\lambda}} > 0$ (constraint) equations $\mathbb{R}^{n_{\boldsymbol{\lambda}}} \ni \boldsymbol{g}(\boldsymbol{q}(t)) = \mathbf{0}$, $n_{\boldsymbol{\lambda}} < n_{\boldsymbol{q}}$.

These constraints restrict the (position) *degrees of freedom* of the system such that the kinematics take place in an $(n_{\boldsymbol{q}} - n_{\boldsymbol{\lambda}})$-dimensional manifold

$$\mathfrak{M}^{\mathrm{s}} := \{\boldsymbol{q} \in \mathbb{R}^{n_{\boldsymbol{q}}} : \boldsymbol{g}(\boldsymbol{q}) = \mathbf{0}\} \subseteq \mathbb{R}^{n_{\boldsymbol{q}}} \ .$$

We will assume that the constraints $\boldsymbol{g} = \mathbf{0}$ are sufficiently smooth such that all terms that appear in the remainder of this section are well-defined and bounded, in particular $\boldsymbol{g}$ should be twice continuously differentiable.

The main idea in the physical modeling process is to simply add terms or, more specifically, forces to the right hand side of (2.2) that 'push' the components of the mechanical system in a way to keep them on the manifold $\mathfrak{M}^{\mathrm{s}}$. Here, the nomenclature 's' refers to the *s*tiff mechanical systems to be introduced in Chapter 3. From physical intuition it is a reasonable requisite that these newly introduced forces do not corrupt the energy behavior of the system. In order not to introduce any work, which would imply the imposition or withdrawal of energy, those forces must always remain orthogonal to the movements. This *D'Alembert's principle* is mathematically equivalent to the introduction of an *extended Lagrangian* (or the classical Lagrangian in the context of constrained minimization problems)

$$\mathcal{L} \rightsquigarrow \mathcal{L} + \boldsymbol{g}(\boldsymbol{q}(t))^{\top}\boldsymbol{\lambda}(t)\,. \tag{2.11}$$

For later reference, we also introduce the *constraint Jacobian*

$$\mathbf{G}(\boldsymbol{q}) := \frac{\partial \boldsymbol{g}(\boldsymbol{q})}{\partial \boldsymbol{q}} \in \mathbb{R}^{n_{\boldsymbol{\lambda}} \times n_{\boldsymbol{q}}}\,, \quad \operatorname{rank} \mathbf{G}(\boldsymbol{q}) = n_{\boldsymbol{\lambda}}\,,$$

where the full-rank (or Grübler-) condition on $\mathbf{G}$ ensures that there are no redundancies or contradictions in the constraint equations.

Formulations of equations of motion Including the constraint forces from (2.11), the equations of motion of a constrained mechanical system (the *index-3 formulation* or *descriptor form* (Luenberger (1977), Brenan et al. (1996)) or *Lagrange equations of the first kind* or *Euler–Lagrange equations*) read

$$\left.\begin{array}{rcl} \mathbf{M}(\boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) & = & \boldsymbol{f}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)) - \mathbf{G}^{\top}(\boldsymbol{q}(t))\boldsymbol{\lambda}(t)\,, \\ \boldsymbol{g}(\boldsymbol{q}(t)) & = & \mathbf{0}\,. \end{array}\right\} \tag{2.12}$$

A justification for the name 'index-3 formulation' will be given in Proposition 2.11. After comparing with (2.5) and using the condensed state vectors $\boldsymbol{y} := (\boldsymbol{q}, \dot{\boldsymbol{q}})^{\top}$, $\boldsymbol{z} := \boldsymbol{\lambda}$ we observe the semi-explicit form of the equations of motion since $\mathbf{M}(\boldsymbol{q})$ is invertible (in a sufficiently large neighborhood of the solution). A differentiation of the constraints leads to the *index-2 formulation*

$$\left.\begin{array}{rcl} \mathbf{M}(\boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) & = & \boldsymbol{f}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)) - \mathbf{G}^{\top}(\boldsymbol{q}(t))\boldsymbol{\lambda}(t)\,, \\ \mathbf{G}(\boldsymbol{q}(t))\dot{\boldsymbol{q}}(t) & = & \mathbf{0}\,. \end{array}\right\} \tag{2.13}$$

Note that the hidden constraints introduce 'another' constraint manifold for the velocities $\dot{\boldsymbol{q}}$, which we call

$$\mathfrak{M}^{\mathrm{d}} := \{(\boldsymbol{q}, \boldsymbol{v}) : \mathbf{G}(\boldsymbol{q})\boldsymbol{v} = \mathbf{0}\} \subseteq \mathbb{R}^{2n_{\boldsymbol{q}}}$$

due to its close relationship to the damped systems. Note that the junction of both, the *tangent bundle*

$$T\,\mathfrak{M}^{\mathrm{s}} := \{(\boldsymbol{q}, \boldsymbol{v}) : \boldsymbol{g}(\boldsymbol{q}) = \mathbf{G}(\boldsymbol{q})\boldsymbol{v} = \mathbf{0}\} \subseteq \mathbb{R}^{2n_{\boldsymbol{q}}}\,,$$

is invariant under the exact flow of (2.12) (presuming consistent initial values). On the other hand the flow of (2.13) only preserves the manifold $\mathfrak{M}^{\mathrm{d}}$ (as long as all terms remain well-defined) and the position constraints $\boldsymbol{g} = \mathbf{0}$, have no relevance in the index-2 case, which we will address in Theorem 2.12 below. Finally, yet another time differentiation of the constraints defines the *index-1 formulation*

$$\left.\begin{array}{rcl} \mathbf{M}(\boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) & = & \boldsymbol{f}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)) - \mathbf{G}^\top(\boldsymbol{q}(t))\boldsymbol{\lambda}(t)\,, \\ \mathbf{G}(\boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) + \mathsf{R}(\boldsymbol{q}(t))(\dot{\boldsymbol{q}}(t), \dot{\boldsymbol{q}}(t)) & = & \mathbf{0}\,, \end{array}\right\} \tag{2.14}$$

where the curvature term (tensor) R has been introduced to collect the second order derivatives of $\boldsymbol{g}$. It satisfies

$$\frac{\partial(\mathbf{G}(\boldsymbol{q})\boldsymbol{u}^{(1)})}{\partial \boldsymbol{q}}\boldsymbol{u}^{(2)} = \mathsf{R}(\boldsymbol{q})(\boldsymbol{u}^{(1)}, \boldsymbol{u}^{(2)}) \quad \text{for all } \boldsymbol{u}^{(1)},\ \boldsymbol{u}^{(2)} \in \mathbb{R}^{n_q}.$$

All three formulations are analytically equivalent (and so equally solvable) if the initial values are consistent to DAE (2.12); yet (2.14) is always well-defined even if the velocity constraints are violated. The acceleration constraints of (2.14) are sometimes difficult and costly to calculate such that many commercial multibody system codes avoid them. Nevertheless, for an exact calculation of consistent initial values (for given $\boldsymbol{q}_0$, $\dot{\boldsymbol{q}}_0$) one has to solve the linear system

$$\begin{pmatrix} \mathbf{M}(\boldsymbol{q}_0) & \mathbf{G}^\top(\boldsymbol{q}_0) \\ \mathbf{G}(\boldsymbol{q}_0) & \mathbf{0} \end{pmatrix} \begin{pmatrix} \ddot{\boldsymbol{q}}(t_0) \\ \boldsymbol{\lambda}(t_0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}(\boldsymbol{q}_0, \dot{\boldsymbol{q}}_0) \\ -\mathsf{R}(\boldsymbol{q}_0)(\dot{\boldsymbol{q}}_0, \dot{\boldsymbol{q}}_0) \end{pmatrix} \tag{2.15}$$

with a *saddle-point structured* matrix that is always non-singular under the above assumptions on $\mathbf{M}$ and $\mathbf{G}$. Commercial codes sometimes employ heuristics to obtain consistent initial values (Leimkuhler et al., 1991, Eich-Soellner and Führer, 1998) but the above saddle-point matrix is often needed in the time integration process anyway. Note that $\mathrm{O}(N)$-formalisms usually work without explicitly forming the matrices $\mathbf{M}$ or $\mathbf{G}$.

Remark 2.10 (Restriction to scleronomic constraints)
In view of Remark 2.3, we also assume that the constraint equations $\boldsymbol{g}(\boldsymbol{q}(t)) = \mathbf{0}$ do not explicitly depend on the time variable t (scleronomic constraints). This restriction is somewhat stronger than just requiring autonomy of an ODE system since the variable t would also enter algebraic equations and hidden constraints.

More precisely, in the ODE case (2.2) *an initial value problem with explicit time dependency in the form*

$$\mathbf{M}(t, \boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) = \boldsymbol{f}(t, \boldsymbol{q}(t), \dot{\boldsymbol{q}}(t))\,, \quad \boldsymbol{q}(t_0) = \boldsymbol{q}_0\,,\ \dot{\boldsymbol{q}}(t_0) = \dot{\boldsymbol{q}}_0,\ t \in [t_0, t_{\text{end}}]$$

can equivalently be transformed into an autonomous equation by treating the time variable t as a dependent variable of $\tau := t$. This extension of the state vector to $\bar{\boldsymbol{q}}(\tau) := (\boldsymbol{q}(\tau), t(\tau))^\top$ as well as adding the trivial equation $t''(\tau) := \mathrm{d}^2 t/(\mathrm{d}\tau^2) = 0$ results in the problem

$$\begin{pmatrix} \mathbf{M}(t(\tau), \boldsymbol{q}(\tau)) & \\ & 1 \end{pmatrix} \bar{\boldsymbol{q}}''(\tau) = \begin{pmatrix} \boldsymbol{f}(t(\tau), \boldsymbol{q}(\tau), \boldsymbol{q}'(\tau)) \\ 0 \end{pmatrix}, \quad \bar{\boldsymbol{q}}(t_0) = \begin{pmatrix} \boldsymbol{q}_0 \\ t_0 \end{pmatrix}, \ \bar{\boldsymbol{q}}'(t_0) = \begin{pmatrix} \dot{\boldsymbol{q}}_0 \\ 1 \end{pmatrix},$$

for $\tau \in [t_0, t_{\text{end}}]$ and with the same analytic solution. However, a simple treatment like this is no longer possible if one deals with non-autonomous and rheonomic constrained mechanical systems (i. e., non-scleronomic constraints $\boldsymbol{g} = \boldsymbol{g}(t, \boldsymbol{q})$ and explicitly time dependent mass matrix $\mathbf{M} = \mathbf{M}(t, \boldsymbol{q})$ and force vector $\boldsymbol{f} = \boldsymbol{f}(t, \boldsymbol{q}, \dot{\boldsymbol{q}})$). In (2.12)*, the matrix function $\mathbf{G}$ is defined as the constraint Jacobian $\partial\boldsymbol{g}/\partial\boldsymbol{q}$. A formal application of the above procedure would result in the DAE system*

$$\begin{aligned} \begin{pmatrix} \mathbf{M}(t(\tau), \boldsymbol{q}(t)) & \\ & 1 \end{pmatrix} \bar{\boldsymbol{q}}''(\tau) = \begin{pmatrix} \boldsymbol{f}(t(\tau), \boldsymbol{q}(\tau), \boldsymbol{q}'(\tau)) \\ 0 \end{pmatrix} - \begin{pmatrix} \mathbf{G}^\top(t(\tau), \boldsymbol{q}(\tau)) \\ \partial\boldsymbol{g}/\partial t \end{pmatrix} \bar{\boldsymbol{\lambda}}(\tau)\,, \\ \boldsymbol{g}(t(\tau), \boldsymbol{q}(\tau)) = \mathbf{0}\,. \end{aligned}$$

Even though the basic mathematical structure (symmetric positive definite matrix on the left hand side, full rank condition of constraint Jacobian) remains the same, additional coupling of $t''(\tau)$ and $\partial \boldsymbol{g}/\partial t$ causes the solutions not to be the same any longer (Arnold, 2016). This problem transfers to the penalty techniques which we will introduce in the next chapter.

Note that we explicitly exclude non-holonomic constraints that depend on the generalized velocities $\dot{\boldsymbol{q}}$ (and cannot be transformed into a purely position-dependent constraint by integration).

In a very broad setting, called 'general mechatronic systems', additional first order equations (e. g. for electronic substructures, thermo-elements or control states)

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\chi}(t, \boldsymbol{x}(t), \boldsymbol{q}(t), \dot{\boldsymbol{q}}(t), \ddot{\boldsymbol{q}}(t), \boldsymbol{\lambda}(t), \boldsymbol{y}(t)) \tag{2.16}$$

are included in the system description, possibly alongside corresponding control and output equations. A comprehensive study of these systems is given by Brüls (2005). For systems with friction models the generalized force vector $\boldsymbol{f}$ might also depend on the Lagrange multipliers $\boldsymbol{\lambda}$ in a nonlinear way and the Grübler condition has to be extended to this setting, too. Last, in its most general form the relation between position coordinates $\boldsymbol{q}$ and velocity variables, here denoted as $\boldsymbol{v} = \dot{\boldsymbol{q}}$, may include another semi-linear relation $\boldsymbol{v}(t) = \boldsymbol{\Omega}(t, \boldsymbol{q}(t))\dot{\boldsymbol{q}}(t)$ such that even the consideration of second order equations embodies a simplification.

Proposition 2.11
The equations of motion in descriptor form are at most of differentiation index three.

Proof. The assertion follows from the observation that (2.12) can be brought into Hessenberg structure (see Example 2.9) using the partitioning of the state vector into $\boldsymbol{z} := \boldsymbol{\lambda}$, $\boldsymbol{y}^{(1)} := \dot{\boldsymbol{q}}$, $\boldsymbol{y}^{(2)} := \boldsymbol{q}$ and defining the new right-hand sides by $\boldsymbol{\varphi}^{(1)} := \mathbf{M}^{-1}(\boldsymbol{f} - \mathbf{G}^\top\boldsymbol{\lambda})$, $\boldsymbol{\varphi}^{(2)} := \dot{\boldsymbol{q}}$, $\boldsymbol{\psi} := \boldsymbol{g}$. The index-3 condition (2.10) transforms to

$$\boldsymbol{\psi}_{\boldsymbol{y}^{(2)}} \cdot \boldsymbol{\varphi}^{(2)}_{\boldsymbol{y}^{(1)}} \cdot \boldsymbol{\varphi}^{(1)}_{\boldsymbol{z}} = \mathbf{G}(\boldsymbol{q}) \cdot \mathbf{I}_{n_{\boldsymbol{q}}} \cdot \mathbf{M}^{-1}(\boldsymbol{q})\mathbf{G}^\top(\boldsymbol{q}) \text{ is invertible,}$$

which is fulfilled as long as the Grübler condition holds and $\mathbf{M}^{-1}$ is well-defined. □

Note that the full-rank condition on $\mathbf{G}$ is sufficient but not necessary for Proposition 2.11. The system is already at most of index three if only the saddle-point matrix in (2.15), for all $\boldsymbol{q}$ and $\dot{\boldsymbol{q}}$, is invertible. We also emphasize that from Proposition 2.11 it follows that the equations of motion (in either formulation) locally have a unique solution.

Using (2.14), the Lagrange multipliers $\boldsymbol{\lambda}(t)$ may theoretically at any point in time and any space vector $(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t))^\top$ be solved for

$$\boldsymbol{\lambda}(t) = \left[(\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^\top)^{-1}\left(\mathbf{G}\mathbf{M}^{-1}\boldsymbol{f} + \mathsf{R}\right)\right](\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)),$$

which can then be inserted into the first set of equations in (2.12). Thus, the solution may be acquired using only two analytic time derivatives of the original equations. Führer and Leimkuhler (1991) call the resulting equations the 'underlying' system of the DAE (2.12) even though it does not coincide with the definition (2.4). Although it is used by many researchers for the integration of multibody systems this approach suffers from a sincere drawback that we are going to discuss in a little more detail:

Drift-off phenomenon A differentiation of the constraints analytically does not affect the solution of the DAE. However, one cannot expect a numerical scheme to do the same since the differentiation leads to a nonlinear change of the state variables in a numerical code. Moreover, and even more importantly, computer methods are subject to discretization and roundoff errors such that the analytic nature of this reformulation gets lost anyway. If we can only assure the estimates $\|\mathbf{G}(\boldsymbol{q}_m)\boldsymbol{v}_m\| \leq c$ for the numerical approximations $\boldsymbol{q}_m$, $\boldsymbol{v}_m$ to $\boldsymbol{q}$, $\dot{\boldsymbol{q}}$ at the mth grid point t_m, ($m = 1, \ldots, n$, see below), $0 < c \ll 1$, we only get an estimation like

$$\|\boldsymbol{g}(\boldsymbol{q}_n)\| \leq \underbrace{\|\boldsymbol{g}(\boldsymbol{q}_0)\|}_{=0} + \int_{t_0}^{t_n} c \,\mathrm{d}t = c \cdot (t_n - t_0)\,,$$

since

$$\boldsymbol{g}(\boldsymbol{q}(t)) = \boldsymbol{g}(\boldsymbol{q}(t_0)) + \int_{t_0}^{t} \Big(\frac{\mathrm{d}}{\mathrm{d}\tau'}\boldsymbol{g}(\boldsymbol{q}(\tau'))\Big)\Big|_{\tau'=\tau} \,\mathrm{d}\tau = \boldsymbol{g}(\boldsymbol{q}_0) + \int_{t_0}^{t} \mathbf{G}(\boldsymbol{q}(\tau))\dot{\boldsymbol{q}}(\tau)\,\mathrm{d}\tau\,,$$

by the Fundamental Theorem of Calculus. In general one can prove the following result.

Theorem 2.12 (Drift-off in mechanical systems (Hairer and Wanner, 2002, Theorem VII.2.1.)) Apply a p-th order convergent numerical method (cf. Definition 2.16 below) to the index-1 formulation (2.14) with consistent initial values. Then the numerical approximations $(\boldsymbol{q}_n, \boldsymbol{v}_n)^\top$ to the solution $(\boldsymbol{q}(t_n), \dot{\boldsymbol{q}}(t_n))^\top$ at $t_n = t_0 + nh$ satisfy on any bounded time interval the estimates

$$\|\boldsymbol{g}(\boldsymbol{q}_n)\| \leq h^p(C_1(t_n - t_0) + C_2(t_n - t_0)^2)\,, \quad \|\mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n\| \leq h^p C_3(t_n - t_0)\,. \tag{2.17}$$

For the index-2 formulation (2.13) the error growth in the position constraints is only linear

$$\|\boldsymbol{g}(\boldsymbol{q}_n)\| \leq C_4 h^p (t_n - t_0) \tag{2.18}$$

with constants $C_i > 0$, $i = 1, \ldots, 4$, that are uniformly bounded and do not depend on the time interval.

Practically, the constants in the error estimations of numerical integrators do indeed depend on the length of the time interval (even exponentially); drift-off is nevertheless a different quality of error since the algebraic equations are actually a part of the given system. Numerical experiments for real-world applications show that the influence of drift-off-induced errors is often a more serious problem than inaccuracy of the method. In the important special case of linear constraints (i. e., constant constraint Jacobian $\mathbf{G}$) also the hidden constraints are linear and—since most time integration methods preserve linear invariants—there is, at least for carefully acquired initial-values, no danger of drift-off.

For the purpose of a thorough overview we also mention that it is possible to solve the equations of motion (2.12) using ODE methods only, within the framework of *local minimal coordinates*: Despite the fact that a global parameterization of the motion by variables of dimension $n_{\boldsymbol{q}} - n_{\boldsymbol{\lambda}}$, i. e., exactly the number of degrees of freedom, so-called *minimal coordinates*, is sometimes impossible and in most cases computationally difficult, one can locally parameterize the constraint manifold $\mathfrak{M}^{\mathrm{s}}$ to get a description of the model that is free of constraints. The work of Wehage and Haug (1982) may be seen as the starting point for the concept of *coordinate partitioning* where a subset of the given coordinates is chosen in each time step such that a complete regular system can be determined.

If we are given a (possibly local) set of minimal coordinates $\boldsymbol{x}(t) \in \mathbb{R}^{n_{\boldsymbol{q}} - n_{\boldsymbol{\lambda}}}$, trivially, the relation

$$\boldsymbol{g}(\boldsymbol{q}(\boldsymbol{x}(t))) = \mathbf{0} \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{g}(\boldsymbol{q}(\boldsymbol{x}(t))) = \mathbf{G}(\boldsymbol{q}(\boldsymbol{x}(t))) \cdot \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{x}} \cdot \dot{\boldsymbol{x}}(t) = \mathbf{0}$$

holds. The conclusio is apparently fulfilled if $\mathbf{N} := \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{x}} \in \mathbb{R}^{(n_{\boldsymbol{q}}-n_{\boldsymbol{\lambda}})\times n_{\boldsymbol{q}}}$ is a nullspace matrix of $\mathbf{G}$, i. e., $\mathbf{G}\cdot\mathbf{N} = \mathbf{0}$ for any argument in a neighborhood of the current state. Here, $\boldsymbol{q}(\boldsymbol{x}(t))$ denotes the dependence of the generalized coordinates on $\boldsymbol{x}$. So, another variant of local minimal coordinates is given by the *tangent space parameterization* (Potra and Rheinboldt, 1991): A linearization of the constraint equations around a point in state-space allows for a local re-parameterization orthogonal to the tangent-bundle $T\,\mathfrak{M}^{\mathrm{s}}$ and so local coordinates with a coordinate transform by the nullspace matrix $\mathbf{N}$. This procedure is at the core of proving the asymptotic expansions for the singular SPPs in Chapter 3 and directly related to the mass-orthogonal projection technique of the next section. In a more abstract and more general setting constrained dynamic equations may always be viewed as *ODEs on manifolds* which also offers specific choices for local minimal coordinates (Rheinboldt, 1984). Either way, a formulation in terms of (local) minimal coordinates leads to the so-called (local) *state space form* (or *Lagrange equations of second kind*) of the multibody system

$$\dot{\boldsymbol{x}}(t) = \bar{\boldsymbol{f}}(\boldsymbol{x}(t)) \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0\,, \tag{2.19}$$

which for later reference is displayed as an explicit first order system. In practical applications, it is often advised to use redundant coordinates $\boldsymbol{q}$ anyway since then the system matrices $\mathbf{M}$ and $\mathbf{G}$ might be sparse leading to smaller computational cost even for larger systems.

Example 2.13 (Minimal coordinates for the mathematical pendulum)
For the example of the point mass under gravitational force in Chapter 1 the kinetic and potential energy are given by

$$\mathcal{T} = \frac{m}{2}((\dot{q}_x(t))^2 + (\dot{q}_y(t))^2)\,, \quad \mathcal{V} = m g_{\mathrm{grav}} q_y(t)\,.$$

Using the geometric relations $q_x(t) = l\sin(\varphi_{\min}(t))$, $q_y(t) = -l\cos(\varphi_{\min}(t))$ the Lagrangian is given by

$$\mathcal{L} = lm\frac{2g_{\mathrm{grav}}\cos(\varphi_{\min}(t)) + l\dot{\varphi}^2_{\min}(t)}{2}\,.$$

Without friction, the above formalism states the differential equation

$$\ddot{\varphi}_{\min}(t) = -\frac{g_{\mathrm{grav}}}{l}\sin(\varphi_{\min}(t))\,.$$

◇

Remark 2.14 (Overdetermined systems, stabilized formulations)
Yet another approach to realize a stable and manageable numerical solution of constrained mechanical systems is to consider the constraint equations on different levels simultaneously (Führer and Leimkuhler, 1991). Because in that case more equations need to be solved for than there are variables, we end up with an overdetermined system *(Campbell, 1987). There are different ways to approach this generic problem class: One is to fix certain equations in the system (usually some of the constraints) and solve the remaining part only in a least squares sense (Barrlund, 1991). Other researchers view an overdetermined system as an approximation problem that can be addressed using methods from nonlinear optimization or inverse problems. A linear combination of the constraints on all three levels is the method suggested by Baumgarte (1972), but it bears the drawback that finding good parameters is a nontrivial task and that artificial high-frequency responses add to the system making the numerical treatment more difficult (Ascher et al., 1994).*

A very common approach in case of multibody system simulation is to add new variables to the system (that should vanish for the exact solution) such that one obtains a system with the same number of equations and unknowns again. The most common way for achieving this is *due to Gear et al. (1985) and called stabilized index-2 formulation or* Gear–Gupta–Leimkuhler formulation*: Here, one takes into account the position and velocity constraints and adds an additional Lagrange multiplier $\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\lambda}}}$ that is then coupled to the trivial relation $\dot{\boldsymbol{q}} = \boldsymbol{v}$ with the velocity $\boldsymbol{v} \in \mathbb{R}^{n_{\boldsymbol{q}}}$:*

$$
\begin{aligned}
\dot{\boldsymbol{q}}(t) &= \boldsymbol{v}(t) - \mathbf{G}^\top(\boldsymbol{q}(t))\boldsymbol{\mu}(t)\,, \\
\mathbf{M}(\boldsymbol{q}(t))\dot{\boldsymbol{v}}(t) &= \boldsymbol{f}(\boldsymbol{q}(t), \boldsymbol{v}(t)) - \mathbf{G}^\top(\boldsymbol{q}(t))\boldsymbol{\lambda}(t)\,, \\
\boldsymbol{g}(\boldsymbol{q}(t)) &= \mathbf{0}\,, \\
\mathbf{G}(\boldsymbol{q}(t))\boldsymbol{v}(t) &= \mathbf{0}\,.
\end{aligned}
$$

The stabilized index-2 formulation may also be seen as a generic way of ensuring the invariant $\mathbf{G}(\boldsymbol{q})\dot{\boldsymbol{q}} = \mathbf{0}$ by a Lagrange multiplier technique (Simeon, 2013). In the literature, the notion of stabilized index-2 formulation is introduced in different ways: Sometimes the velocity relation is multiplied by $\mathbf{M}(\boldsymbol{q})$ (or a lumped version of it), see (Hairer and Wanner, 2002, Sect. VII.1). A practical drawback of Gear–Gupta–Leimkuhler formulation is that often $\mathbf{G}$ is not at hand in the computational realization but there exists techniques using the Jacobian (of the nonlinear systems occuring in the time integration) to approximate it (see Arnold et al., 2011). In the original work Gear et al. (1985) propose to use a staggered procedure to obtain the numerical solution in each time step which may be interpreted as a projection *after one constraint is already fulfilled. As projection techniques in a broader sense are also a common way to realize constraint enforcement and, as we will see in the next chapters, that for initial values away from $T\mathfrak{M}^{\mathrm{s}}$ it is important to have 'the natural way' of finding corresponding values fulfilling the constraints, we will shortly present the most important details on this matter in the next section.*

2.1.3 Projection onto the constraint manifold

Given (poorly chosen) initial values or rough numerical approximations to the state variables $(\bar{\boldsymbol{q}}, \bar{\boldsymbol{v}}) \approx (\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t))$ one is faced with the problem of finding a genuine way to relate them to certain values fulfilling the constraints. To simplify notations we start with the introduction of the Delassus matrix (Brogliato, 2013):

$$
\mathbf{S}(\boldsymbol{q}) := \Big[\mathbf{G}\mathbf{M}^{-1}\mathbf{G}^\top\Big](\boldsymbol{q})\,. \tag{2.20}
$$

Note that since $\mathbf{M}$ is symmetric and positive definite and we assume the Grübler condition, $\mathbf{S}$ is always well-defined and non-singular. A (nonlinear) projection onto $T\mathfrak{M}^{\mathrm{s}}$ may now be accomplished using the following mappings.

Mass-orthogonal projection: Define

$$
\boldsymbol{q} = \boldsymbol{\pi}(\bar{\boldsymbol{q}}) := \bar{\boldsymbol{q}} - \Big[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\Big](\boldsymbol{q})\,\boldsymbol{\nu} \quad \text{such that } \boldsymbol{g}(\boldsymbol{\pi}(\bar{\boldsymbol{q}})) = \mathbf{0}\,. \tag{2.21}
$$

for the position variables $\bar{\boldsymbol{q}}$ and (afterwards)

$$
\boldsymbol{v} = \mathbf{P}\bar{\boldsymbol{v}} \quad \text{with} \quad \mathbf{P} := \mathbf{I} - \Big[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\mathbf{G}\Big](\boldsymbol{q}) \in \mathbb{R}^{n_{\boldsymbol{q}} \times n_{\boldsymbol{q}}} \tag{2.22}
$$

for the velocity variables.

Remark 2.15 (Well-definition and interpretation of the projection map)

(a) *The definition of $\boldsymbol{\pi}$ is implicit: Formally, to obtain the projected values one has to solve the system of nonlinear equations*

$$
\mathbf{0} = \boldsymbol{\Psi}_0(\boldsymbol{q}, \boldsymbol{\nu}; \bar{\boldsymbol{q}}) := \begin{pmatrix} \boldsymbol{q} - \big(\bar{\boldsymbol{q}} - \big[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\big](\boldsymbol{q})\boldsymbol{\nu}\big) \\ \boldsymbol{g}(\boldsymbol{q}) \end{pmatrix}\,, \tag{2.23}
$$

which is locally uniquely solvable within a neighborhood of $(\bar{\boldsymbol{q}}, \mathbf{0})^\top$ *as long as* $\|\boldsymbol{g}(\bar{\boldsymbol{q}})\| \ll 1$ *because* $\boldsymbol{\Psi}_0(\bar{\boldsymbol{q}}, \boldsymbol{\nu}; \bar{\boldsymbol{q}}) = (\mathbf{0}, \boldsymbol{g}(\bar{\boldsymbol{q}}))^\top$ *and the Jacobian*

$$\begin{aligned}\frac{\partial \boldsymbol{\Psi}_0(\boldsymbol{q}, \boldsymbol{\nu}; \bar{\boldsymbol{q}})}{\partial(\boldsymbol{q}, \boldsymbol{\nu})} &= \begin{pmatrix} \mathbf{I} & \left[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\right](\boldsymbol{q}) \\ \mathbf{G}(\boldsymbol{q}) & \mathbf{0} \end{pmatrix} + \mathcal{O}(1) \cdot \|\boldsymbol{\nu}\| \\ &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{G}(\boldsymbol{q}) & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \left[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\right](\boldsymbol{q}) \\ \mathbf{0} & -\mathbf{I} \end{pmatrix} + \mathcal{O}(1) \cdot \|\boldsymbol{\nu}\|\end{aligned}$$

is non-singular as it can be decomposed into the product of regular matrices. Well-definition of $\boldsymbol{\pi}$ *in a sufficiently small neighborhood of* $\mathfrak{M}^{\mathrm{s}}$ *thus follows from the Implicit Function Theorem. Note that one might also take a different argument* $\tilde{\boldsymbol{q}} \in \mathfrak{M}^{\mathrm{s}}$ *of* $[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}]$ *in (2.21). As long as* $\tilde{\boldsymbol{q}}$ *and* $\boldsymbol{q}$ *(or* $\tilde{\boldsymbol{q}}$ *and* $\bar{\boldsymbol{q}}$ *respectively) are sufficiently close to each other it is still guaranteed that the projection is well-defined. In the analysis in Chapter 5 below we will use both cases: Once the implicit nonlinear projection* $\boldsymbol{\pi}$ *has been applied to find consistent initial values for a constrained mechanical system (2.12) one can use the analytic solution* $\tilde{\boldsymbol{q}} := \boldsymbol{q}(t)$ *to define the projection.*

(b) *Matrix* $\mathbf{P}$ *is always well-defined and a projector onto the tangential space* $T_{\boldsymbol{q}}\mathfrak{M}^{\mathrm{s}} = \ker \mathbf{G}(\boldsymbol{q})$, *because* $\mathbf{P}\boldsymbol{v} = \boldsymbol{v}$ *if and only if* $\boldsymbol{v} \in T_{\boldsymbol{q}}\mathfrak{M}^{\mathrm{s}}$ *and* $\mathbf{P}^2 = \mathbf{P}$, *as can easily be verified. Note that*

$$\mathbf{P} \cdot [\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}) = \mathbf{0}\,. \tag{2.24}$$

(c) *From a theoretical viewpoint, the definition of the nonlinear projection* $\boldsymbol{\pi}$ *might also be stated as a minimization problem: If we presume the solution* $\boldsymbol{q}^* := \boldsymbol{\pi}(\bar{\boldsymbol{q}})$ *known a-priori it is solution of the constrained minimization problem*

$$\min_{\boldsymbol{q} \in \mathbb{R}^{n_{\boldsymbol{q}}}} \|\bar{\boldsymbol{q}} - \boldsymbol{q}\|_*\,, \;\; s.\,t.\;\; \boldsymbol{g}(\boldsymbol{q}) = \mathbf{0}\,, \tag{2.25}$$

where the norm $\|\bullet\|_* := \sqrt{(\bullet)^\top \mathbf{M}(\boldsymbol{q}^*)(\bullet)}$ *is mass-matrix induced. If, on the other hand, we consider the above case where the argument* $\boldsymbol{q}$ *in (2.23) is to be replaced by* $\tilde{\boldsymbol{q}} \in \mathfrak{M}^{\mathrm{s}}$ *the norm in (2.25) is induced by* $\mathbf{M}(\tilde{\boldsymbol{q}})$.

For the definition of $\mathbf{P}$ *one equivalently minimizes for given* $\bar{\boldsymbol{v}} \in \mathbb{R}^{n_{\boldsymbol{q}}}$ *the norm*

$$\min_{\boldsymbol{v} \in \mathbb{R}^{n_{\boldsymbol{q}}}} \|\bar{\boldsymbol{v}} - \boldsymbol{v}\|_*\,, \;\; s.\,t.\;\; \mathbf{G}(\boldsymbol{q})\boldsymbol{v} = \mathbf{0}$$

with linear *constraints, such that the solution is given by the linear relation in (2.22). In the literature this form of projection is therefore often denoted as* **M**-orthogonal *or* mass-orthogonal projection. *From that point of view the variable* $\boldsymbol{\nu}$ *in the definition of* $\boldsymbol{\pi}$ *is just another Lagrange multiplier from the constrained minimization problem (2.25) and the entire construction of* $\boldsymbol{\pi}$ *and* $\mathbf{P}$ *an explicit solution of the necessary condition for stationarity. Note that these necessary conditions have again the saddle-point matrix from (2.15) at their core. From physical as well as mathematical point of view this definition is the most natural choice because (i) it is invariant under affine transformations of the coordinates (Lubich, 1991) and (ii) as the norm is mass-induced it is a minimal-work approximation: If we view the projection as the physical process of moving the bodies to a consistent position, this one requires the least energy.*

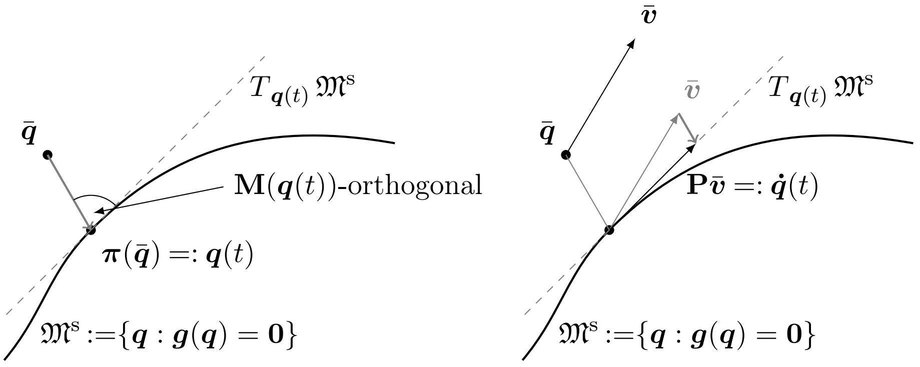

$T_{\boldsymbol{q}(t)}\,\mathfrak{M}^{\mathrm{s}}$
$\bar{\boldsymbol{q}}$
$\mathbf{M}(\boldsymbol{q}(t))$-orthogonal
$\boldsymbol{\pi}(\bar{\boldsymbol{q}}) =: \boldsymbol{q}(t)$
$\mathfrak{M}^{\mathrm{s}} := \{\boldsymbol{q} : \boldsymbol{g}(\boldsymbol{q}) = \mathbf{0}\}$
$\bar{\boldsymbol{v}}$
$\bar{v}$
$T_{\boldsymbol{q}(t)}\,\mathfrak{M}^{\mathrm{s}}$
$\bar{\boldsymbol{q}}$
$\mathbf{P}\bar{\boldsymbol{v}} =: \dot{\boldsymbol{q}}(t)$
$\mathfrak{M}^{\mathrm{s}} := \{\boldsymbol{q} : \boldsymbol{g}(\boldsymbol{q}) = \mathbf{0}\}$

Figure 2.1: Projections and projectors to the constraint manifold

2.2 Numerical methods

After this overview on the theoretical basis of physical multibody system modeling in technical simulation we will now turn our attention to the numerical solution of the equations of motion. To keep the presentation compact we will mainly concentrate on the ODE case. So, for first order systems we refer to the state space form (2.19). We also exclude any detailed software aspects as, for instance, step size strategies. Since the time integration methods of Newmark-type are going to be introduced in detail in Chapter 4 this will only be a very basic summary for an easier reference in the next chapters. For a more detailed discussion of the methods and original literature we refer to the monographs of Hairer et al. (1993) and Hairer and Wanner (2002).

The very basic idea of almost all time integration methods for initial value problems is to simply follow the flow of the ODE/DAE: Taking the initial value $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$ as a first 'approximation' we advance forward in time in $N > 0$ steps of length $h := (t_{\text{end}} - t_0)/N$ to acquire numerical approximations $\boldsymbol{x}_n \approx \boldsymbol{x}(t_n)$, $t_n = t_0 + nh$, $n = 1, \ldots, N$, one after another. To value the accuracy of the method the following two fundamental concepts of consistency and convergence are necessary.

Definition 2.16 (Consistency, Convergence)**.**

(a) A numerical scheme has *order of consistency* $p \geq 0$ if there exists a constant $h_0 > 0$ such that the numerical solution after one step for all $h \in (0, h_0]$ fulfills the estimate

$$\|\boldsymbol{x}_1 - \boldsymbol{x}(t_1)\| \leq C h^{p+1}\,,$$

where $C > 0$ is a (problem dependent, bounded) constant that does not depend on the time step size h. If $p \geq 1$, the method is *consistent*.

(b) If for all $h \in (0, h_0]$ the approximations fulfill the estimate

$$\|\boldsymbol{x}_n - \boldsymbol{x}(t_n)\| < \tilde{C} h^p, \;\; (\tilde{C} > 0,\; n = 0, 1 \ldots, N) \tag{2.26}$$

the method has *order (of convergence) p*. If (2.26) holds for any $p > 0$ the method is said to be *convergent*.

For some methods, so-called *multistep methods* more than one approximation at a previous step is necessary. For the definition of consistency in that case we consider the solution for all $t \leq t_0$ to be exactly given (and extend $\boldsymbol{x}(t)$ to this time interval if necessary). So, for this family of methods the order of convergence may depend on the initialization. For some more complex schemes one may obtain different orders for different components of the state vector as, for example, in multibody dynamics the Lagrange multipliers are defined on acceleration level. If the order of the method drops when applied to certain problem classes (beyond the classical setting) one speaks of *order reduction.* The usual assumption on the problem involves a Lipschitz condition on $\bar{\boldsymbol{f}}$ with respect to the Picard–Lindelöf Theorem that ensures unique solvability of the system. To obtain high-order results $\bar{\boldsymbol{f}}$ is typically assumed to have smooth bounded derivatives to a certain order.

2.2.1 Time integration schemes in technical simulation

The design of numerical schemes for problems of multibody dynamics is subject to characteristic demands of this problem class: As more complex models may take into account impact problems, unilateral constraints and discontinuous control states and rely on data base look-ups and interpolations for input values, the systems fall short of the strong smoothness assumptions needed for high order methods. Additionally, there are uncertainties in the parameters and the external inputs, and sometimes insufficient mathematical models (cf. Example 2.6). Moreover, for multiphysics problems the order (in space and time) of the method is bounded inasmuch as there is an additional spatial error, i. e., an error of the space discretzation (see Example 3.19 below) and yet another error source from the coupling of sub-systems in co-simulation applications.

Due to a theorem of Dahlquist (1963) the order of so-called *unconditionally stable* multistep methods is bounded by two and, especially in structural dynamics applications, the actual computational models use local linearizations of $\bar{\boldsymbol{f}}$ (Hoff and Pahl, 1988b), resulting in an order reduction to at most order two for most schemes. Eberhard and Schiehlen (1998) propose a hierarchical development of models where components of the system are modeled at different levels of accuracy and detail such that accuracy is less an issue than a reliable rough approximation of the solution. In conclusion, for applications in multibody dynamics often it is not useful to design high order methods as one cannot benefit from their superior convergence properties and it is recommended to use robust and efficient methods of order $p = 1, 2$. It might also seem rather rudimentary that in this thesis we only consider a fixed time step size h, but with regards to real time and large scale applications this reflects the current state-of-the-art in some branches.

Onestep methods The first family of numerical procedures for initial value problems for ODEs we will turn our attention to are *(implicit) Runge–Kutta methods.* Applied to second order systems (2.2) and with the initialization $\boldsymbol{q}_0 := \boldsymbol{q}(t_0)$, $\boldsymbol{v}_0 := \dot{\boldsymbol{q}}(t_0)$ one step is described by

$$\begin{aligned} \boldsymbol{Q}_n^{(i)} &= \boldsymbol{q}_n + h \textstyle\sum_{j=1}^s a_{ij} \dot{\boldsymbol{Q}}_n^{(j)} , \quad & \dot{\boldsymbol{Q}}_n^{(i)} &= \boldsymbol{v}_n + h \textstyle\sum_{j=1}^s a_{ij} \ddot{\boldsymbol{Q}}_n^{(j)} , \quad (i = 1, \ldots, s) , \\ \boldsymbol{q}_{n+1} &= \boldsymbol{q}_n + h \textstyle\sum_{i=1}^s b_i \dot{\boldsymbol{Q}}_n^{(i)} , \quad & \boldsymbol{v}_{n+1} &= \boldsymbol{v}_n + h \textstyle\sum_{i=1}^s b_i \ddot{\boldsymbol{Q}}_n^{(i)} , \end{aligned} \tag{2.27}$$

where the stage vectors $(\boldsymbol{Q}_n^{(i)}, \dot{\boldsymbol{Q}}_n^{(i)}, \ddot{\boldsymbol{Q}}_n^{(i)})^\top$ satisfy the *equilibrium condition*

$$\mathbf{M}(\boldsymbol{Q}_n^{(i)}) \ddot{\boldsymbol{Q}}_n^{(i)} = \boldsymbol{f}(\boldsymbol{Q}_n^{(i)}, \dot{\boldsymbol{Q}}_n^{(i)}) .$$

(For this discussion we disregard explicit Runge–Kutta methods, i. e. methods (2.27) with $a_{ij} = 0$, $j < i$, due to their inferior stability properties.) For later reference we state that the *stability function* R: $\mathbb{C} \to \mathbb{C}$ of the Runge–Kutta method with the Runge–Kutta matrix $\mathbf{A} := (a_{ij})_{i,j=1,\ldots,s} \in \mathbb{R}^{s \times s}$ and the weight vector $\boldsymbol{b} := (b_j)_{j=1,\ldots,s} \in \mathbb{R}^s$ is defined as

$$R(z) := \frac{\det(\mathbf{I}_s - z\mathbf{A} + z\mathbb{1}\boldsymbol{b}^\top)}{\det(\mathbf{I} - z\mathbf{A})} ,$$

where $\mathbf{I}_s$ denotes the identity matrix in $\mathbb{R}^{s\times s}$ and $\mathbb{1} := (1,\dots,1)^\top \in \mathbb{R}^s$. The *stability region* S of a Runge–Kutta method is the set of all $z \in \mathbb{C}$ with $|R(z)| \leq 1$; if $\{z : \Re(z) \leq 0\} \subseteq S$, the method is A-stable; if additionally $\lim_{z\to\infty} R(z) = 0$, it is called L-stable; I-stability is given if $\mathbb{R}\cdot \mathrm{i} \subseteq S$. The coefficients of the methods are usually summarized in a so-called *Butcher tableau*

$$\begin{array}{c|c} \boldsymbol{c} & \mathbf{A} \\ \hline & \boldsymbol{b}^\top \end{array}\,,$$

where the *stage vector* $\boldsymbol{c} \in \mathbb{R}^s$ is defined by $c_i := \sum_{j=1}^s a_{ij}$, $i = 1,\dots,s$. From (2.27) we see that the stage vectors $\boldsymbol{Q}_n^{(i)}$ serve as approximations to $\boldsymbol{q}(t_n + c_i h)$ such that a more detailed notion of consistency–stage order–is helpful here. In short, stage order $p_\mathrm{s} \geq 0$ implies that for the application to quadrature problems the Runge-Kutta method's stage vectors $\dot{\boldsymbol{Q}}_n^{(j)}$, $\ddot{\boldsymbol{Q}}_n^{(j)}$, $j = 1,\dots,s$, are order p_s approximations to the solution. This property can also be expressed in terms of the coefficients:

Definition 2.17 (Stage order)**.**
The Runge–Kutta method with coefficients $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c})$ has *stage order* $p_\mathrm{s} \geq 0$ if

$$\sum_{j=1}^{s} a_{ij} c_j^{k-1} = \frac{c_i^k}{k} \quad \text{for } i = 1,\dots,s,\ k = 1,\dots,p_\mathrm{s}\,.$$

The most commonly used (implicit) Runge–Kutta methods are Radau-, Gauss- and Lobatto methods that originate from the corresponding quadrature rules and are favored for their high (stage) order and stability properties. In view of the rather low accuracy requirements in technical simulation we just state that the (second order) trapezoidal rule falls into the family of Lobatto-IIIA methods, the implicit midpoint-rule is a one-stage method of Gauss type and the (only first order) implicit Euler scheme is a Radau-IIA method. The same is true for the algorithm RADAU5 (Hairer and Wanner, 2002) which is a fifth order implicit Runge–Kutta code that is very often used as a reference for numerical experiments.

Especially higher order implicit Runge–Kutta methods have the drawback of high computational cost as all stage vectors are coupled and so each step involves the solution of a very large nonlinear system. There are several approaches known from the literature to remedy this problem: *Diagonally implicit* (DIRK) methods allow for a staggered procedure: The stage vectors are computed consecutively such that at least the dimension of the systems is lowered. For *Rosenbrock* and *Rosenbrock–Wanner methods*, on the other hand, the nonlinear systems are linearized such that each step requires only the solution of a large *linear* system.

Linear multistep methods A methodology that allows for high order and small nonlinear systems is given by linear multistep methods. Instead of computing stage vectors in each time step one re-uses approximations at previous time steps. This comes at the cost of additional work for initialization, more complex code for variable step size implementations and worse stability properties. If a multistep method employs $k > 1$ states from previous time steps, it is called k-step method. These methods are also preferable because the equilibrium condition is at the new time instance t_{n+1} which is beneficial for DAEs since then the constraints are always exactly fulfilled. For Runge–Kutta methods this requires so-called stiff accuracy. In the class of linear multistep methods backward-differentiation formulae, *BDF methods* (also called Gear's methods) are most commonly used in the context of DAE simulation. As their name suggests they are based on backward finite difference approximations of the first derivative. The BDF 'one-step' method coincides with the implicit Euler scheme and the two-stage BDF scheme, in short BDF(2), is given by

$$\tfrac{3}{2}\boldsymbol{x}_{n+1} - 2\boldsymbol{x}_n + \tfrac{1}{2}\boldsymbol{x}_{n-1} = h\bar{\boldsymbol{f}}(\boldsymbol{x}_{n+1})\,. \tag{2.28}$$

Methods for second order systems The relation between the time derivative of position coordinates and velocities is so simple that it seems rather artificial to consider this equation explicitly as a part of the differential equation. As also ODEs and DAEs of second order are historically very important whenever a mechanically motivated model needs to be solved, there is a large variety of methods tailored to mechanical systems. Newmark methods in the generalized form of Chung and Hulbert (1993) are the main concern of this thesis; so we postpone their introduction to Chapter 4 below. Within the class of linear multistep methods there are second-derivative multistep methods (of Enright) and second-order BDF-type methods (see Hairer and Wanner, 2002, Sect. V.3) but they are not that commonly used in the multibody dynamics community. An exception might be methods of Verlet-type and some specialized schemes for Hamiltonian problems, but they lack the strong stability properties needed for many large-scale or very stiff problems in engineering.

In the family of onestep methods the concept of partitioned Runge–Kutta methods is at the basis for the construction of specialized methods for second order systems (Runge–Kutta–Nyström methods): The left- and right-hand side of (2.27) then involve *different* Runge–Kutta parameters $\mathbf{A}$, $\boldsymbol{b}$ and so provide more degrees of freedom when optimizing the parameters to specific needs.

Using the same idea within the context of linear multistep methods goes back to the work of Dahlquist (1959). With the representation given by Console and Hairer (2014), *partitioned linear multistep methods* for the coupled system

$$\begin{aligned}\dot{\boldsymbol{q}}(t) &= \boldsymbol{v}(t)\,,\\ \dot{\boldsymbol{v}}(t) &= \boldsymbol{f}(t,\boldsymbol{q},\boldsymbol{v})\end{aligned}$$

can be expressed as

$$\begin{aligned}\sum_{i=0}^{k} \alpha_i^{\boldsymbol{q}} \boldsymbol{q}_{n+i-k+1} &= h \sum_{i=0}^{k} \beta_i^{\boldsymbol{q}} \boldsymbol{v}_{n+i-k+1}\,,\\ \sum_{i=0}^{k} \alpha_i^{\boldsymbol{v}} \boldsymbol{v}_{n+i-k+1} &= h \sum_{i=0}^{k} \beta_i^{\boldsymbol{v}} \boldsymbol{f}(t_{n+i-k+1}, \boldsymbol{q}_{n+i-k+1}, \boldsymbol{v}_{n+i-k+1})\end{aligned} \tag{2.29}$$

for the *two* parameter sets $(\alpha_i^{\boldsymbol{q}}, \beta_i^{\boldsymbol{q}})_{i=0,\dots,k}$, $(\alpha_i^{\boldsymbol{v}}, \beta_i^{\boldsymbol{v}})_{i=0,\dots,k}$ with $\alpha_k^{\boldsymbol{q}}$, $\alpha_k^{\boldsymbol{v}} \neq 0$. In Chapter 4 we will see that Newmark time integrators fall into that framework of (generalized) linear multistep methods.

2.2.2 Stiffness and strongly attractive systems

In the above sections we already used the term 'stiff' to describe that a problem has a characteristic challenging nature concerning its numerical treatment. A strict definition of stiffness is very difficult as it has to take into account that stiffness may depend on dimension, initial value, smoothness and may even vary throughout time. We will use the very pragmatic and historically first explanation given by Curtiss and Hirschfelder (1952, not explicitly stated in the reference) that for stiff systems certain implicit solvers, in particular BDF, perform tremendously better than explicit ones.

There is a quasi-consensus on the terminology when *dissipative systems* are solved. Yet, there is still active research on a mathematically profound definition of 'stiffness' (Söderlind et al., 2015). Some researchers, see the above reference, argue that stiff mechanical systems (where the stiffness is of physical nature and not to be confused with the numerical term stiffness) are not to be considered as stiff since they are too close to being ill-conditioned.

Dissipative systems are characterized by their fundamentally different behavior forward (attractive, stable) or backward (repulsive, unstable) in time. The strongly attractive mechanical systems that are one main subject of this thesis fall into this spectrum: As they describe how external impacts on the mechanical system are damped out, a time inversion leads to an exponential growth of even smallest excitations. From their nature they are strongly connected to and sometimes used for the consistent initialization of DAEs (Brenan et al., 1996, Leimkuhler et al., 1991) as will become clear in Theorem 3.8. Problems of a similar mathematical structure also appear in chemical reaction kinetics where certain reactions happen orders of magnitude faster than others. The spectrum of the Jacobian of the right-hand side for this type of problem is exactly on or very close to the negative real axis which allows for certain specialized techniques for their numerical solution. In particular, within certain limitations, it is even possible to design explicit methods with many stages that are still stable for these problems (Medovikov, 1998, Abdulle, 2002).

2.2.3 Highly oscillatory systems

The numerical treatment of problems with high frequency oscillations is probably one of the most difficult but also most important tasks when any continuous phenomena in the natural sciences need to be analyzed or simulated by means of numerical procedures (Petzold et al., 1997). Prior to the selection of an appropriate algorithm one is always confronted with certain questions on what one expects it to achieve. If it is indeed essential to resolve the oscillations to acquire useful information from the model it is in most cases quite unavoidable to put large computational effort into the calculations. Hughes (1987) and Cardona and Géradin (1994) advise to use approximately ten points to render one period of oscillation sufficiently. For the stiff mechanical systems of the next section this requirement is not even worth discussing.

For applications in multibody system dynamics a fine resolution of high frequency oscillations is never the aim of the engineer. On the one hand the reason for that lies in the lower accuracy requirements (see Section 2.2.1). On the other hand one should not forget that the simulation run is usually just a small fraction of the design process. Mostly, the results are further processed as input for optimization algorithms or as control variables of other components of a complex system. For control theory applications non-oscillating inputs are usually favored due to stability issues even if that assumption is provably wrong (Siciliano and Book, 1988).

Since highly oscillatory systems are an immense field of research we give only an enumeration of the most important approaches. This section is based on the review articles (Petzold et al., 1997, Cohen et al., 2006, Abdulle et al., 2012) where more detailed information and further references can be found.

The challenging character of highly oscillatory problems obligates specialized solutions for the problem at hand. There is no, and probably will not ever be, an all-purpose method for oscillatory problems. As the goal usually is a stable integration using large time steps the methods are often referred to as ‘long-time-step methods’. A classification of those methods may be based on typical properties of the problem as

(a) one constant high frequency in the model,

(b) one almost constant high frequency that is (i) time dependent or (ii) state dependent,

(c) weak coupling of oscillatory and non-oscillatory components or

(d) linear highly oscillatory terms,

(e) Hamiltonian systems allowing to take advantage of structural properties (symmetries, time-reversibility, adiabatic invariants/effective Hamiltonians), among others. A very general framework is the *heterogeneous multiscale method* (HMM) (E and Engquist, 2003) that covers problems from (a)-(d). For the application of HMM to highly-oscillatory problems one typically adapts methods from the classical theory of averaging methods (Arnold, 1988, Sanders et al., 2007) to construct a so-called macroscopic model that is easier to tackle. This procedure is carried out using small time steps to locally resolve the model accurately and then average the right-hand sides to evolve in time with a superimposed second method, called macro-solver. In the context of the stiff mechanical systems of this thesis, those methods have been applied by Ariel et al. (2012). A special case tailored for the use in mechanical system simulation is the mollified impulse method that is analyzed in detail by Calvo and Sanz-Serna (2009), see also the analysis of a projected analogon by Lubich and Weiss (2014). Within this framework of numerical averaging techniques (somewhat a limit case) fall *stroboscopic methods* (Minorsky, 1962, Petzold, 1981) that are based on a discrete sampling of the right hand side at *exactly* the same phase but are restricted to methods from problem classes (a) and (b).

Problems from (c) and (d) are often approached using general *multiscale* or *multirate* techniques. A starting point for the mathematical analysis is the work of Gear and Wells (1984). For Hamiltonian problems (e) there has been large progress in the construction of so-called *variational integrators* (Lew et al., 2004). These methods follow the 'discretize first' approach: Instead of applying numerical schemes to the necessary conditions of the Hamiltonian principle the energy or Lagrangian of the model itself is discretized and the system solved using a discrete variational principle. Within this area fall *energy-momentum methods* (Simo and Tarnow, 1992) where the focus is on preservation of symmetries of the system, e. g. angular momentum or energy.

At last, there are also many *regularization* approaches that act on the equations before a numerical scheme is applied. For the problems considered in this work *quasistatic approaches* (Jahnke et al., 1993) are naturally of interest: Sometimes it is computationally cheaper to consider the limit case of infinite stiffness (infinite frequencies) leading to additional constraints in the system. An established way is Guyan (also: Irons–Guyan) reduction for linear systems that is mathematically based on Schur complements of the system matrices. Model reduction techniques as Craig–Bampton or static condensation fall into this branch as well (Hughes, 1987). The exactly opposite approach is also used in practical simulations: For a large class of systems it is possible to define a cut-off frequency $f_0 > 0$ and regulate all higher modes to coincide with f_0.

The concern of this thesis is to evaluate how far it is possible to rely on the given equations and Newmark integration methods while ensuring a stable and accurate numerical solution *measured with respect to the limit of infinite stiffness.* In the next chapter we establish the analytic foundations concerning the two 'stiff' problem classes in mechanical system simulation.

Chapter 3

Singularly perturbed systems in multibody dynamics

This chapter is devoted to the analysis of the mathematical structure, inherent in the two substitute problems from Chapter 1. In particular, we will see that for very large spring or damping constants, both physically motivated models may be characterized as *singularly perturbed problems* (SPPs). We will start by formally introducing the broad concept of SPPs, highlight the degenerated class of *singular* SPPs and then come back to the physical modeling of mechanical systems by means of singular force terms.

The introduction of this chapter is to some extend based on the monograph of O'Malley Jr. (1991). The analysis of analytic properties of the mechanical systems with singular force terms is mainly based on work of Lubich (1993) and Stumpp (2008). In this work, we will only consider perturbation problems for ODEs, recognizing that a series of problems for partial differential equations (PDEs) may be studied using similar techniques. At any rate, from a computational viewpoint, we already cover the important case of large ODE systems stemming from finite-element or finite-difference discretizations, see Example 3.19 below and (Simeon, 2013, Altmann, 2015) for a comprehensive discussion on the relations of DAE and PDE-models in technical simulation.

Assumption 3.1. From now on we will always assume that there is a sufficiently small but positive constant $\varepsilon_0 > 0$, depending on the problem under consideration, such that the parameter ε (resp. δ in Section 3.2 below) can be estimated by ε_0:

$$0 < \varepsilon, \delta < \varepsilon_0$$

3.1 Singularly perturbed systems

The field of perturbation analysis deals with dynamic problems, i. e., mainly differential equations, depending on small parameters. So, methods from this field always come into play when there are negligibly small parameters or unproportionally large ones, respectively, or more than one characteristic scale, in time or space, is present. Often, SPPs represent a *regularization* (Bornemann, 1998) of a given unperturbed problem and are therefore artificially introduced.

To understand the essence of SPPs, at first we consider the *regular* perturbation problem

$$\ddot{q}^\varepsilon(t) + \varepsilon q^\varepsilon(t) = 0\,, \quad q^\varepsilon(0) = q_0,\ \ \dot{q}^\varepsilon(0) = \dot{q}_0 \tag{3.1}$$

with a small (perturbation-) parameter $0 < \varepsilon \ll 1$ and the exact solution

$$q^\varepsilon(t) = \sqrt{q_0^2 + \varepsilon^{-1}\dot{q}_0^2}\cos(\sqrt{\varepsilon}t + \varphi_0)\,, \quad \varphi_0 = \operatorname{atan2}(-\dot{q}_0, \sqrt{\varepsilon}q_0)\,,$$

where we assume that atan2(x, y) coincides with $\arctan(x/y)$ for x, $y \in \mathbb{R}$, $y > 0$.

If ε approaches zero, the solution $q^\varepsilon(t)$ tends *uniformly on any finite time interval* towards the solution of $\ddot{q}^0(t) = 0$, $q^0(0) = q_0$, $\dot{q}^0(0) = \dot{q}_0$, i. e., the system that is obtained if one formally sets ε to zero. This result is *independent* of the specific values of q_0, $\dot{q}_0$ and the higher order derivatives of $q^\varepsilon(t)$ are bounded independently of ε.

If, on the other hand, we consider the coefficient of $q^\varepsilon(t)$ in (3.1) (3.1) to become very large replacing ε by ε^{-1}, we get after rescaling the problem

$$\dot{q}^\varepsilon(t) = v^\varepsilon(t)\,, \quad \varepsilon\dot{v}^\varepsilon(t) + q^\varepsilon(t) = 0\,, \quad q^\varepsilon(0) = q_0\,, \quad \dot{q}^\varepsilon(0) = \dot{q}_0\,. \tag{3.2}$$

Its solution shows a highly oscillatory behavior as $\varepsilon \to 0$. In particular, no upper bound on the derivatives of $q^\varepsilon(t)$ can be determined unless the initial values are chosen to be exactly zero. Indeed, when formally plugging in $\varepsilon = 0$, we no longer deal with a differential equation but instead have an algebraic relation that determines $q^0(0) \equiv 0$ and leaves no degrees of freedom for the initial values. We will see below that (3.2) is already the degenerate case of a *singular* SPP but the main features are apparent: For $\varepsilon \to 0$ there is no uniform convergence towards a smooth function and for $\varepsilon = 0$ the equation degenerates.

In a more general way, and remembering (2.5), we state the problem class as

$$\begin{aligned} \dot{\boldsymbol{y}}^\varepsilon(t) &= \boldsymbol{\varphi}(\boldsymbol{y}^\varepsilon(t), \boldsymbol{z}^\varepsilon(t); \varepsilon)\,, \\ \varepsilon\dot{\boldsymbol{z}}^\varepsilon(t) &= \boldsymbol{\psi}(\boldsymbol{y}^\varepsilon(t), \boldsymbol{z}^\varepsilon(t); \varepsilon) \end{aligned} \tag{3.3}$$

for $\boldsymbol{y}^\varepsilon \in \mathbb{R}^{n_{\boldsymbol{y}}}$, $\boldsymbol{z}^\varepsilon \in \mathbb{R}^{n_{\boldsymbol{z}}}$, $t \in [t_0, t_{\text{end}}]$, $t_0 < t_{\text{end}}$, and given initial values at t_0. In the literature $\boldsymbol{y}^\varepsilon$ are usually called the *slow* or *smooth variables* whereas $\boldsymbol{z}^\varepsilon$ are denoted as *fast* or *sharp variables.* The *reduced problem* is obtained if ε is formally set to zero

$$\begin{aligned} \dot{\boldsymbol{y}}^0(t) &= \boldsymbol{\varphi}(\boldsymbol{y}^0(t), \boldsymbol{z}^0(t); 0)\,, \\ \boldsymbol{0} &= \boldsymbol{\psi}(\boldsymbol{y}^0(t), \boldsymbol{z}^0(t); 0)\,. \end{aligned} \tag{3.4}$$

As we have already seen, one cannot expect the solutions of (3.3) always to converge uniformly to solutions of the reduced system (3.4) which follows simply from a dimension argument: Initial values of the reduced problem are constrained to fulfill (at least) $\boldsymbol{\psi} = \boldsymbol{0}$ whereas for the original problem the choice is, or appears to be, free of any additional conditions. The goal of perturbation theory lies in finding a way to express the limiting behavior of the systems in a very general way. Practically this is done by searching for series expansions

$$\boldsymbol{y}^\varepsilon(t) = \underbrace{\boldsymbol{y}^0(t) + \sum_{i=1}^{\infty} \varepsilon^i \boldsymbol{y}^i(t)}_{=:\boldsymbol{y}^{\text{sm}}(t)} + \boldsymbol{y}_{\text{bl}}(t)\,, \qquad \boldsymbol{z}^\varepsilon(t) = \underbrace{\boldsymbol{z}^0(t) + \sum_{i=1}^{\infty} \varepsilon^i \boldsymbol{z}^i(t)}_{=:\boldsymbol{z}^{\text{sm}}(t)} + \boldsymbol{z}_{\text{bl}}(t)\,,$$

where $(\boldsymbol{y}^{\text{sm}}(t), \boldsymbol{z}^{\text{sm}}(t))^\top$ is called the *outer expansion* and $(\boldsymbol{y}^{\text{bl}}(t), \boldsymbol{z}^{\text{bl}}(t))^\top$ is called *boundary layer.* If the coefficients $\boldsymbol{y}^i$, $\boldsymbol{z}^i$, $i = 1, 2, \ldots$, of the outer expansion are bounded on $[t_0, t_{\text{end}}]$ we call $(\boldsymbol{y}^{\text{sm}}(t), \boldsymbol{z}^{\text{sm}}(t))^\top$ *smooth expansion.* The crucial part of the analysis of SPPs is usually the estimation of the boundary layer. For stable regular SPPs, see Definition 3.2 below, one can show that it is negligible apart from a small region near t_0. For later reference, we call $(\boldsymbol{y}^0(t), \boldsymbol{z}^0(t))^\top$ the *slow solution* or, with respect to (2.5) the *DAE solution.* In the literature, smooth or slow solutions are sometimes also called averaged solution. As this is technically not always justified we use the upper nomenclature. For an analysis of the relationship between averaged and smooth motion in the context of stiff mechanical systems, see (Reich, 1995, Ariel et al., 2012, Brumm and Weiss, 2014).

Definition 3.2 (Singularly perturbed system (SPP)).
The system (3.3) with $n_{\boldsymbol{y}} \geq 0$, $n_{\boldsymbol{z}} \geq 1$ is called *(regular) singularly perturbed system* if the reduced problem (3.4) has a solution $(\boldsymbol{y}^0(t), \boldsymbol{z}^0(t))^\top$ and the Jacobian $\mathbf{J} := \partial\boldsymbol{\psi}/\partial\boldsymbol{z}^\varepsilon$ is nonsingular in a sufficiently small neighborhood of $(\boldsymbol{y}^0(t), \boldsymbol{z}^0(t))^\top$. If, additionally, there exists a $\beta > 0$ such that in that neighborhood all eigenvalues λ_i, $i = 1, 2, \ldots, n_{\boldsymbol{z}}$, of $\mathbf{J}$ fulfill

$$\Re(\lambda_i) \leq -\beta\,,$$

the SPP is called *stable*.

Note that by (2.8), Definition 3.2 implies that the reduced problem is of index one. Having a look at (3.2), and interpreting q^ε as the slow, v^ε as the fast variable, we see that $\mathbf{J} \equiv 0$ is obviously singular. As we will see in Chapter 4, (3.2) is not only an exemplary representative for equations in mechanical system analysis but is in fact *the* standard problem in the investigation of numerical time integration schemes. So, there is need to broaden the scope of problems under consideration.

Definition 3.3 (Singular singularly perturbed system).
We will call an SPP *singular* if for $\varepsilon > 0$ for the reduced problem (3.4) a differential-algebraic system of index two or higher is attained. If the limiting problem is of index one it is called *regular* (or just singularly perturbed system).

Note that, to keep the representation compact, for both definitions we neglected the initial values which may have an effect on the class and, more importantly, have to be taken into account to define what is meant by 'the solution' of the reduced problem. We will always assume that the initial values of the SPPs are sufficiently close to consistent initial values of the reduced problem where in Assumption 5.25 below we will fix what we mean by 'sufficiently' in the context of singularly perturbed mechanical systems. Note also that for nonlinear systems the character (singular or regular SPP) may change throughout time evolution. This leads to the theory of so-called *shock layers* which lies beyond the scope of this work.

Remark 3.4 (Alternative characterization of singular SPPs)
In textbooks on SPPs (O'Malley Jr., 1991, Shchepakina et al., 2014) the characterization of singular SPPs follows the somewhat more vague definition given by Flaherty and O'Malley Jr. (1980). Here, singular SPPs are characterized by the fact that solutions of the reduced problem locally define a nontrivial manifold. So, in this more general sense, the class of singular SPPs may even include problems without a well-defined differentiation index at all. To ensure well-posedness there are usually additional contractivity assumptions on the problem imposed or explicit nonlinear transformations constructed such that well-definition of the slow motion is guaranteed.

In the above definition we follow Becker et al. (2014), see also the introduction to Chap. VI in (Hairer and Wanner, 2002), p. 452 in (Petzold et al., 1997) and (Gu et al., 1989). Other researchers refer to this problem class as SPPs 'in the critical case' motivated by the work of Vasil'eva and Butuzov (1980) or 'non-standard' SPPs (Etchechoury and Muravchik, 2003). Note also that sometimes regular (but non-stable) SPPs are also considered as singular ones.

Remark 3.5 (Singularly perturbed DAEs)
As mechanical multibody systems in their general form are often subject to constraint equations, the consideration of just the ODE case falls short of giving a comprehensive description. To give a more generic definition one can also consider the perturbed problem class of DAE-type

$$\begin{aligned} \dot{\boldsymbol{y}}^\varepsilon(t) &= \boldsymbol{\varphi}(\boldsymbol{y}^\varepsilon(t), \boldsymbol{z}^\varepsilon(t), \varepsilon)\,, \\ \mathbf{0} &= \bar{\boldsymbol{\psi}}(\boldsymbol{y}^\varepsilon(t), \boldsymbol{z}^\varepsilon(t), \dot{\boldsymbol{z}}^\varepsilon(t), \varepsilon) \end{aligned}$$

as has been done by Yan (1997) and Higueras (2001). As analytic and numerical properties of index-1 DAEs are very similar to those of ODEs and many techniques and results can be transferred to this case, we will also widen Definition 3.3 in the sense that if for $\varepsilon > 0$ the problem already is an index-1 DAE, we still use the term 'regular SPP'.

A general extension of the classical theory of SPPs to singularly perturbed DAEs is difficult as the singular character—non-uniform or even no convergence towards the reduced problem—can occur even if $\partial\bar{\psi}/\partial\dot{\boldsymbol{z}} = \mathbf{I}_{n_{\boldsymbol{z}}}$. As a result, the literature on the subject is rather sparse and mostly concentrated on special cases. Rheinboldt and Simeon (1999) propose a problem class structured like

$$\begin{aligned}\begin{pmatrix}\mathbf{M}_{\text{rigid}}(\boldsymbol{q}_{\text{slow}},\boldsymbol{q}_{\text{fast}}) & \mathbf{C}^\top(\boldsymbol{q}_{\text{slow}},\boldsymbol{q}_{\text{fast}})\\ \mathbf{C}(\boldsymbol{q}_{\text{slow}},\boldsymbol{q}_{\text{fast}}) & \mathbf{M}_\Delta\end{pmatrix}\begin{pmatrix}\ddot{\boldsymbol{q}}_{\text{slow}}\\ \ddot{\boldsymbol{q}}_{\text{fast}}\end{pmatrix} & \\ = \begin{pmatrix}\boldsymbol{f}_{\text{rigid}}(\boldsymbol{q}_{\text{slow}},\boldsymbol{q}_{\text{fast}},\dot{\boldsymbol{q}}_{\text{slow}},\dot{\boldsymbol{q}}_{\text{fast}})\,,\\ \boldsymbol{f}_\Delta(\boldsymbol{q}_{\text{slow}},\boldsymbol{q}_{\text{fast}},\dot{\boldsymbol{q}}_{\text{slow}},\dot{\boldsymbol{q}}_{\text{fast}}) - \nabla\frac{1}{\varepsilon^2}\,\mathcal{U}(\boldsymbol{q}_{\text{fast}})\end{pmatrix} & \\ - \mathbf{G}^\top(\boldsymbol{q}_{\text{slow}},\boldsymbol{q}_{\text{fast}})\boldsymbol{\lambda}\,, & \\ \boldsymbol{g}(\boldsymbol{q}_{\text{slow}},\boldsymbol{q}_{\text{fast}}) = \mathbf{0}\,. &\end{aligned} \tag{3.5}$$

which can be seen as a DAE extension of the stiff mechanical systems we are going to review in detail in Section 3.3. Typically, for being able to carry out the analysis, one imposes a transversality condition as stated by Simeon (2013, p. 177) or Bornemann (1998, page 21, Definition 3): The manifold defined via the (hard) constraints $\boldsymbol{g} = \mathbf{0}$ and the manifold stemming from the (weak) constraints imposed by the stiff potential $\frac{1}{\varepsilon^2}\,\mathcal{U}(\boldsymbol{q}_{\text{fast}})$ should intersect in a 'non-flat' (Bornemann, 1998) way. Roughly speaking, this simply implies that $\boldsymbol{g} = \mathbf{0}$ and $\nabla\mathcal{U} = \mathbf{0}$ do not contradict one another or coincide. In the original work, Rheinboldt and Simeon (1999) restrict the analysis to linear stiff force terms such that the transversality condition may be verified more easily. Note the strong connection between the transversality condition and the existence of a reparameterization in terms of local minimal coordinates, preserving the structure of the mechanical systems. Practically, (3.5) *is very important since it represents flexible multibody systems in their most general form (Simeon, 2013).*

Yen and Petzold (1998) are also concerned with highly oscillatory DAEs with SPP character but their analysis is mainly guided by computational considerations. Weber et al. (2012) propose quasistatic approaches, i. e., generic ways of computationally obtaining the limiting system for $\varepsilon \to 0$ *without explicitly deriving its equations.*

In the next two sections we introduce two prototypical problem classes that appear in mechanical system simulation. Both are derived starting from the principle that a certain type of constraints is supposed to be approximately conserved by the model via the introduction of singular force terms. Notice that from now on we disregard the DAE case or 'hard constraints' and assume that, without the singular forces, the system may always be described by an ODE (2.2).

Judging from the great improvement for numerical methods conserving given quantities or invariants of systems, this may seem rather artificial and as if the problems are made more difficult than necessary using this approach. But one should always keep in mind: (a) Sometimes the mathematical structure of the two problem classes is somewhat hidden and the involved quantities, i. e., primarily $\boldsymbol{g}$, $\mathbf{G}$ and the exact value of the perturbation parameter, are not known or difficult to acquire. (b) Also, modeling mechanical systems using 'hard constraints' is itself always an abstraction from the physical world or as van Kampen and Lodder (1984) put it: 'The constraints of classical mechanics are [always] idealizations of stiff springs'. Judging from the great progress of parallel computing in recent years, it should also be noted that the replacement of joints by spring-dampers is in fact easily parallelizable, and that data-exchange in a complex simulation environment might be easier if the coordinates have a straightforward geometric interpretation.

3.2 Strongly damped mechanical systems

The problem class, to which we refer here as strongly damped mechanical systems is a generalization of the damped pendulum from Chapter 1 not only with regards to the mathematical structure but also from the viewpoint of mechanical modeling: To derive the equations of motion of strongly damped mechanical systems, we make use of a Rayleigh dissipation function that penalizes if the generalized velocities of the mechanical system violate some hidden substitutes of holonomic constraints. An archetypical application case is the simulation of biomechanical systems with articular surfaces or cartilage tissue by means of soft constraints as proposed by Hans (2004).

Following the framework of a Rayleigh dissipation function from Example 2.6, defining the state dependent dissipation function

$$\mathcal{D} := \frac{1}{2\delta} \left\| \mathbf{G}(\boldsymbol{q}^\delta(t))\dot{\boldsymbol{q}}^\delta(t) \right\|_2^2 ,$$

and $\delta \cdot \frac{\partial \mathcal{D}}{\partial \dot{\boldsymbol{q}}^\delta} = [\mathbf{G}^\top \mathbf{G}](\boldsymbol{q}^\delta)\dot{\boldsymbol{q}}^\delta$, we attain the system

$$\mathbf{M}(\boldsymbol{q}^\delta(t))\ddot{\boldsymbol{q}}^\delta(t) = \boldsymbol{f}(\boldsymbol{q}^\delta(t), \dot{\boldsymbol{q}}^\delta(t)) - \frac{1}{\delta}\mathbf{G}^\top(\boldsymbol{q}^\delta(t))\mathbf{G}(\boldsymbol{q}^\delta(t))\dot{\boldsymbol{q}}^\delta(t) , \tag{3.6}$$

where instead of ε we used $\delta > 0$ as perturbation parameter to distinguish from the stiff mechanical systems in Section 3.3 below and for easier reference. With the partitioning $\boldsymbol{y}^\delta := \boldsymbol{q}^\delta$, $\boldsymbol{z}^\delta := \boldsymbol{v}^\delta := \dot{\boldsymbol{q}}^\delta$, and recalling that $\mathbf{M}(\boldsymbol{q}^\delta)$ is symmetric positive definite, this system is of the form (3.3) and we have

$$\frac{\partial \boldsymbol{\psi}(\boldsymbol{y}^\delta, \boldsymbol{z}^\delta)}{\partial \boldsymbol{z}^\delta} = \frac{\partial \left([\mathbf{M}^{-1}\mathbf{G}^\top \mathbf{G}](\boldsymbol{q}^\delta)\dot{\boldsymbol{q}}^\delta \right)}{\partial \dot{\boldsymbol{q}}^\delta} + \mathcal{O}(\delta) = [\mathbf{M}^{-1}\mathbf{G}^\top \mathbf{G}](\boldsymbol{q}^\delta) + \mathcal{O}(\delta) \in \mathbb{R}^{n_q \times n_q} ,$$

which is clearly rank-deficient for $\delta \to 0$ such that the system is no regular SPP. If instead we formally introduce the 'multiplier-like' variables $\boldsymbol{\lambda}^\delta := \frac{1}{\delta}\mathbf{G}(\boldsymbol{q}^\delta)\dot{\boldsymbol{q}}^\delta$, (3.6) may equivalently be stated as

$$\mathbf{M}(\boldsymbol{q}^\delta(t))\ddot{\boldsymbol{q}}^\delta(t) = \boldsymbol{f}(\boldsymbol{q}^\delta(t), \dot{\boldsymbol{q}}^\delta(t)) - \mathbf{G}^\top(\boldsymbol{q}^\delta(t))\boldsymbol{\lambda}^\delta(t) , \tag{3.7a}$$

$$\delta \boldsymbol{\lambda}^\delta(t) = \mathbf{G}(\boldsymbol{q}^\delta(t))\dot{\boldsymbol{q}}^\delta(t) , \tag{3.7b}$$

which we will call the 'index-1 formulation' of (3.6). That (3.7) is in fact of index one can be verified by time differentiation of (3.7b) which yields

$$\dot{\boldsymbol{\lambda}}^\delta(t) = \frac{1}{\delta}\mathbf{G}(\boldsymbol{q}^\delta(t))\ddot{\boldsymbol{q}}^\delta(t) + \mathsf{R}(\boldsymbol{q}^\delta(t))(\dot{\boldsymbol{q}}^\delta(t), \dot{\boldsymbol{q}}^\delta(t)) ,$$

or by considering (3.7) as a Hessenberg system with constraint $\mathbf{0} = \boldsymbol{\psi} := \mathbf{G}(\boldsymbol{q}^\delta)\dot{\boldsymbol{q}}^\delta - \delta\boldsymbol{\lambda}^\delta$, and invertible constraint Jacobian $-\delta \cdot \mathbf{I}_{n_{\boldsymbol{\lambda}}}$.

Note that the modeling process includes the introduction of a *dissipation* function. So—by construction—the analytic solution of the system for $\delta > 0$ loses energy if we assume a conservative system in absence of the singular force terms. As a consequence of Corollary 3.9 below, we will nevertheless see that in the limit case $\delta \to 0$ the solution approaches the DAE-solution in index-2 formulation and so the energy loss/energy error vanishes as well. A formal insertion of $\delta = 0$ in (3.7) shows that (2.13) in fact describes the corresponding slow motion. As a consequence, strongly damped mechanical systems are singular SPPs.

Remark 3.6 (Stable behavior of solutions)
From the viewpoint of analytic stability one may argue that (3.6) *is not singular in the sense that it shows an unstable behavior or a degenerate boundary layer response. It is, on the contrary, possible (Stumpp, 2008, Lemma 4) to transform* (3.6) *to the form*

$$\begin{aligned} \dot{\boldsymbol{y}}^\delta(t) &= \tilde{\boldsymbol{\varphi}}(\boldsymbol{y}^\delta(t), \boldsymbol{z}^\delta(t), \delta)\,, \\ \delta\dot{\boldsymbol{z}}^\delta(t) &= \mathbf{W}(\boldsymbol{y}^\delta(t))\boldsymbol{z}^\delta(t) + \delta\bar{\boldsymbol{\psi}}(\boldsymbol{y}^\delta(t), \boldsymbol{z}^\delta(t), \delta)\,, \end{aligned} \tag{3.8}$$

with a function $\tilde{\boldsymbol{\varphi}}\colon \mathbb{R}^{n_{\boldsymbol{q}}+n_{\boldsymbol{\lambda}}} \times \mathbb{R}^{n_{\boldsymbol{q}}-n_{\boldsymbol{\lambda}}} \to \mathbb{R}^{n_{\boldsymbol{q}}+n_{\boldsymbol{\lambda}}}$ *and a symmetric and positive definite matrix* $\mathbf{W}\colon \mathbb{R}^{n_{\boldsymbol{q}}-n_{\boldsymbol{\lambda}}} \to \mathbb{R}^{n_{\boldsymbol{\lambda}}\times n_{\boldsymbol{\lambda}}}$. *So, in view of Definition 3.2, we are dealing with a regular SPP after a nonlinear coordinate transformation. Etchechoury and Muravchik (2003) present a generic way and conditions such that such a transformation to a regular (and stable) SPP is always possible. In conclusion, we see that Definition 3.3 depends on the specific choice of coordinates and may change using analytic transformations. Judging from the fact that we used the differentiation index for the definition, this comes without surprise since analytic transformations, differentiations, are the core of index-reduction. Kramer (2006) studies linear multistep methods (BDF) for the solution of SPPs that may be lipeomorphically transformed to a stable SPP and calls these systems* quasi singularly perturbed. *(Recall that a lipeomorphism is a homomorphism which is Lipschitz continuous and has a Lipschitz continuous inverse.)*

System (3.8) is not the only nonlinear coordinate transform that simplifies the situation: The following lemma shows that, for the purpose of analysis, it suffices to consider linear damping terms and a clear separation of the velocity coordinates.

Lemma 3.7 (Strongly damped mechanical systems: Alternative formulation (Stumpp, 2008, Lemma 3))
For given $(\boldsymbol{q}^0, \boldsymbol{v}^0)^\top \in \mathfrak{M}^{\mathrm{d}}$ there exists, locally but independent of $\delta > 0$, a smooth coordinate change $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{q})$, $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{q}, \dot{\boldsymbol{q}})$, such that in the new coordinates the Rayleigh dissipation function may be written as

$$\mathcal{D} = \mathcal{D}(\boldsymbol{z}) = \frac{1}{2\delta}\|\boldsymbol{z}^\perp\|_2^2\,,$$

where $\boldsymbol{z} = (\boldsymbol{z}^\parallel, \boldsymbol{z}^\perp)^\top$ is partitioned in components $\boldsymbol{z}^\parallel \in \mathbb{R}^{n_{\boldsymbol{q}}-n_{\boldsymbol{\lambda}}}$ in direction of velocities 'consistent' with $\mathfrak{M}^{\mathrm{d}}$ and $\boldsymbol{z}^\perp \in \mathbb{R}^{n_{\boldsymbol{\lambda}}}$ in normal direction. In the transformed variables, the strongly damped system (3.6) reads

$$\begin{aligned} \mathbf{H}(\boldsymbol{y}(t))\dot{\boldsymbol{y}}(t) &= \boldsymbol{z}(t)\,, \\ \tilde{\mathbf{M}}(\boldsymbol{y}(t))\dot{\boldsymbol{z}}(t) &= \tilde{\boldsymbol{f}}(\boldsymbol{y}(t), \boldsymbol{z}(t)) - \frac{1}{\delta}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_{\boldsymbol{\lambda}}} \end{pmatrix} \boldsymbol{z}(t) \end{aligned}$$

with $\tilde{\mathbf{M}} := \mathbf{H}^{-\top}(\boldsymbol{y})\mathbf{M}(\boldsymbol{y})\mathbf{H}^{-1}(\boldsymbol{y})$ and $\tilde{\boldsymbol{f}} := \mathbf{H}^{-\top}\left(\boldsymbol{f}(\boldsymbol{y}, \mathbf{H}^{-1}(\boldsymbol{y})\boldsymbol{z}) - \mathbf{M}(\boldsymbol{y})\frac{\partial \mathbf{H}^{-1}(\boldsymbol{y})}{\partial \boldsymbol{y}}(\mathbf{H}^{-1}(\boldsymbol{y})\boldsymbol{z}, \boldsymbol{z})\right)$. The matrix $\mathbf{H} \in \mathbb{R}^{n_{\boldsymbol{q}}\times n_{\boldsymbol{q}}}$ is implicitly defined: Let

$$[\mathbf{G}^\top\mathbf{G}](\boldsymbol{q}) =: \mathbf{Q}^\top(\boldsymbol{q}) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [\mathbf{L}\mathbf{L}^\top](\boldsymbol{q}) \end{pmatrix} \mathbf{Q}(\boldsymbol{q}) \tag{3.9}$$

be a separation of $\mathbf{G}^\top\mathbf{G}$'s singular and regular parts with an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{n_{\boldsymbol{q}}\times n_{\boldsymbol{q}}}$. The $(n_{\boldsymbol{\lambda}} \times n_{\boldsymbol{\lambda}})$-block $[\mathbf{L}\mathbf{L}^\top]$ is symmetric and regular and may again be decomposed into its Cholesky factors $\mathbf{L}$. Matrix $\mathbf{H}$ is then constructed as

$$\mathbf{H}(\boldsymbol{q}) := \begin{pmatrix} \mathbf{I}_{n_{\boldsymbol{q}}-n_{\boldsymbol{\lambda}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}^\top \end{pmatrix} \mathbf{Q}(\boldsymbol{q})\,.$$

To avoid confusion, we omitted the superscript δ in this lemma. Note that a splitting like in (3.9) is only possible if the full-rank condition on $\mathbf{G}$ remains fulfilled. The limiting behavior of (3.6) can be described as follows.

Theorem 3.8 (Smooth motion, invariant manifold (Stumpp, 2004, 2008))
For arbitrary fixed $i_{\max} \geq 1$ and each pair $(\boldsymbol{q}_0^0, \dot{\boldsymbol{q}}_0^0)^\top \in \mathbb{R}^{2n_{\boldsymbol{q}}}$ satisfying

$$\mathbf{G}(\boldsymbol{q}_0^0)\dot{\boldsymbol{q}}_0^0 = \mathbf{0} \in \mathbb{R}^{n_{\boldsymbol{\lambda}}} ,$$

there exists a pair $(\boldsymbol{q}_0^\delta, \dot{\boldsymbol{q}}_0^\delta)^\top$, unique to order $\mathcal{O}(\delta^{i_{\max}})$, and with

$$\dot{\boldsymbol{q}}_0^0 - \dot{\boldsymbol{q}}_0^\delta \in \mathcal{O}(\delta) \cap \left(T_{\boldsymbol{q}_0^0}\,\mathfrak{M}^{\mathrm{d}}\right)^{\mathbf{M}(\boldsymbol{q}_0^0)-\perp} \tag{3.10}$$

such that the solution $(\boldsymbol{q}^\delta(t), \dot{\boldsymbol{q}}^\delta(t))^\top$ of (3.6) with initial values $(\boldsymbol{q}_0^\delta, \dot{\boldsymbol{q}}_0^\delta)^\top$ is smooth in the sense that it has bounded derivatives to order $i_{\max}$ and allows for an expansion of the form

$$\begin{aligned} \boldsymbol{q}^\delta(t) &= \boldsymbol{q}^0(t) + \delta\boldsymbol{q}^1(t) + \ldots + \delta^{i_{\max}}\boldsymbol{q}^{i_{\max}}(t) + \mathcal{O}(\delta^{i_{\max}+1}) , \\ \dot{\boldsymbol{q}}^\delta(t) &= \dot{\boldsymbol{q}}^0(t) + \delta\dot{\boldsymbol{q}}^1(t) + \ldots + \delta^{i_{\max}}\dot{\boldsymbol{q}}^{i_{\max}}(t) + \mathcal{O}(\delta^{N+1}) , \end{aligned} \tag{3.11}$$

where $\boldsymbol{q}^i(t)$, $i \geq 0$, are independent of $\delta > 0$ and exist on finite time intervals. In particular, $\boldsymbol{q}^0(t) = \boldsymbol{q}(t)$, $t \in [t_0, t_{\mathrm{end}}]$, is the solution of the index-2 equations of motion of the corresponding constrained system. All values $(\boldsymbol{q}_0^\delta, \dot{\boldsymbol{q}}_0^\delta)^\top$ form a $(2n_{\boldsymbol{q}} - n_{\boldsymbol{\lambda}})$-dimensional manifold $\mathfrak{M}^\delta \subset \mathbb{R}^{2n_{\boldsymbol{q}}}$, which is invariant under the flow of (3.6) to terms in $\mathcal{O}(\delta^{i_{\max}+1})$.

Corollary 3.9 (Rubin–Ungar Theorem (I))
The *well-defined* solutions of the equations of motion of strongly damped mechanical systems (3.6) with initial values $\boldsymbol{q}^\delta(t_0) = \boldsymbol{q}(t_0)$, $\dot{\boldsymbol{q}}^\delta(t_0) = \dot{\boldsymbol{q}}(t_0)$ that are consistent to the DAE system (2.12), uniformly on any finite time interval $[t_0, t_{\mathrm{end}}]$, approach the solutions of the equations of motion in their index-2 form (2.13). The differences of position and velocity coordinates remain in $\mathcal{O}(\delta)$. The result remains true if the deviation of $(\boldsymbol{q}_0^\delta, \dot{\boldsymbol{q}}_0^\delta)^\top$ from $\mathfrak{M}^{\mathrm{d}}$ is $\mathcal{O}(\delta)$.

Note that the consistency with the position constraints $\boldsymbol{g}(\boldsymbol{q}_0^\delta) = \mathbf{0}$ is imposed to guarantee that the solution to the DAE exists, i. e., that all involved values are always well-defined. If we only required consistency with the velocity constraint $\mathbf{G}(\boldsymbol{q}^\delta(t_0))\dot{\boldsymbol{q}}^\delta(t_0) = \mathbf{0}$ we would possibly face a severe drift-off from $\boldsymbol{g} = \mathbf{0}$ as in the index-2 case in (2.13) such that the arguments might not remain within the domain of $\boldsymbol{f}$, $\mathbf{M}$ or $\mathbf{G}$ respectively. Note also that we adapted the original statement of the theorem from Stumpp (2008) to indicate the correspondence to the stiff mechanical systems below. Equation (3.10) implies that only projection at velocity level, i. e., using $\mathbf{P}$ is necessary, cf. Figure 3.1.

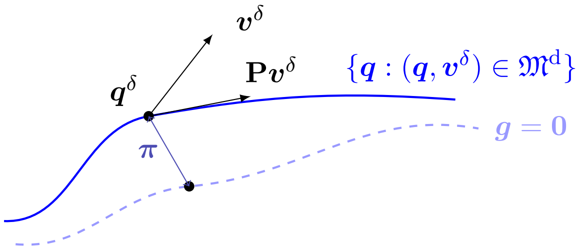

$\boldsymbol{v}^\delta$
$\mathbf{P}\boldsymbol{v}^\delta$
$\{\boldsymbol{q} : (\boldsymbol{q}, \boldsymbol{v}^\delta) \in \mathfrak{M}^{\mathrm{d}}\}$
$\boldsymbol{q}^\delta$
$\boldsymbol{g} = \mathbf{0}$
π

Figure 3.1: Schematic illustration of projections and projectors in (2.21) and (2.22)

Example 3.10 (Attractive invariant manifold for two test problems)

(a) *Inclined plane*
If, instead of a circle, the motion of the point mass in the pendulum example (1.1) is constrained to the plane $0 = g(q_x, q_y) = q_x - q_y$, the equations of motion describe a pointmass sliding without any friction on an inclined surface. Taking initial conditions $q_x(0) = q_y(0) = \dot{q}_x(0) = \dot{q}_y(0) = 0$, the solution can be calculated analytically:

$$q_{x|y}(t) = -\frac{g_{\text{grav}}}{4} t^2 \,, \quad t \geq 0 \,,$$

which corresponds to the slow motion component $\boldsymbol{q}(t)$. For the strongly attractive systems, and taking for simplicity unit constants $g_{\text{grav}} = m = 1$, we get the system

$$\begin{aligned} \ddot{q}_x^\delta(t) &= \tfrac{1}{\delta}(\dot{q}_y^\delta(t) - \dot{q}_x^\delta(t)) \\ \ddot{q}_y^\delta(t) &= -1 + \tfrac{1}{\delta}(\dot{q}_x^\delta(t) - \dot{q}_y^\delta(t)) \,. \end{aligned}$$

With the initial values $(q_x^\delta(0), q_y^\delta(0)) = (q_x^0, q_y^0)$, $(\dot{q}_x^\delta(0), \dot{q}_y^\delta(0)) = (\dot{q}_x^0, \dot{q}_y^0)$ this system can once again be solved analytically:

$$\begin{aligned} q_x^\delta(t) &= -\frac{t^2}{4} + \frac{1}{8}\left(-\delta^2 + 2\delta(\dot{q}_x^0 - \dot{q}_y^0 + t) + 4(2q_x^0 + \dot{q}_x^0 t + \dot{q}_y^0 t)\right) + \frac{1}{8}\delta\left(\delta - 2\dot{q}_x^0 + 2\dot{q}_y^0\right) \mathrm{e}^{-\frac{2t}{\delta}} \,, \\ q_y^\delta(t) &= \underbrace{-\frac{t^2}{4} + \frac{1}{8}\left(\delta^2 - 2\delta(\dot{q}_x^0 + \dot{q}_y^0 - t) + 4(2q_y^0 + \dot{q}_x^0 t + \dot{q}_y^0 t)\right)}_{\text{smooth motion}} \; \underbrace{-\frac{1}{8}\delta\left(\delta - 2\dot{q}_x^0 + 2\dot{q}_y^0\right) \mathrm{e}^{-\frac{2t}{\delta}}}_{\text{boundary layer}} \,. \end{aligned}$$

Clearly, the nonsmooth part, i. e., the boundary layer solution component, vanishes if the condition on the initial velocities

$$\frac{\delta}{2} = \dot{q}_x^0 - \dot{q}_y^0 \tag{3.12}$$

holds. That means that all higher order coefficients in the series expansion of Theorem 3.8 vanish for this linear example. In addition, we note the following two observations:

(i) (3.12) does not exactly describe the constraint equation on velocity level $\mathbf{G}(\boldsymbol{q})\dot{\boldsymbol{q}} = \mathbf{0}$ but instead a $\mathcal{O}(\delta)$ deviation from it.

(ii) Inserting these analytic solutions into the original constraint equations $0 = q_x - q_y$ reveals a linear drift-off:

$$g(q_x^\delta(t), q_y^\delta(t)) = 2\delta t \,.$$

This is the general situation for singular SPPs with fast and slow variables and an attractive invariant manifold (Nipp, 2002).

Notice that the boundary layer components are independent of the position variables $(q_x, q_y)^\top$. This is due to $\mathbf{G}$ being constant which is an important special case as we will see in Example 3.19.

(b) *Mathematical pendulum*
For the pendulum equations (1.1) no analytic solution can be determined. Nevertheless, it is possible to equate the initial values of the coefficients in (3.11) to get an arbitrarily close approximation to the smooth motion. We consider the initial values of the DAE system as in Chapter 1. To obtain smooth initial values for (1.3), the series expansion (3.11) is inserted into (3.7) and all involved quantities are expanded into Taylor series with respect to δ.

Equating the coefficients shows that $(\boldsymbol{q}^i(t), \dot{\boldsymbol{q}}^i(t))^\top$, $i = 0, 1, \ldots$, may be defined recursively as solutions to index-2 DAEs of the form (2.13), once the preceding coefficients $(\boldsymbol{q}^j(t), \dot{\boldsymbol{q}}^j(t))^\top$, $j = 0, 1, \ldots, i-1$, and their time derivatives are considered as known quantities. As for the computation of consistent initial values in (2.15), this leads to a generic way to compute the initial values to arbitrary order in terms of powers of δ. The additional condition (3.10) is necessary for uniqueness of the initial values. For the chosen setting we get for $i = 1, 2, \ldots$ the additional equations $\dot{q}_x^i(t_0) = \dot{q}_y^i(t_0)$. Note, however, that the computation of higher order derivatives may require very involved computations. In fact, obtaining the solution $(\boldsymbol{q}^i(t), \dot{\boldsymbol{q}}^i(t))^\top$ without prior knowledge of preceding coefficients involves the solution of an index-$(2+2k)$ DAE system. Note also that $\mathfrak{M}^\delta$ imposes no restrictions on the position coordinates, i. e., $(q_x^i(t_0), q_y^i(t_0))^\top$, $i = 1, 2, \ldots$, may be chosen freely. With respect to approximating the true pendulum motion, it is of course reasonable to let them simply vanish. For the first four series coefficients, we obtain

$$\dot{q}_x^1(0) = \dot{q}_y^1(0) = \sqrt{2} - \tfrac{g_{\mathrm{grav}}}{2}\,, \qquad \dot{q}_x^2(0) = \dot{q}_y^2(0) = \tfrac{3g_{\mathrm{grav}}}{\sqrt{2}}\,,$$
$$\dot{q}_x^3(0) = \dot{q}_y^3(0) = \tfrac{3(8+g_{\mathrm{grav}}^2)}{2\sqrt{2}}\,, \qquad \dot{q}_x^4(0) = \dot{q}_y^4(0) = 27\sqrt{2}g_{\mathrm{grav}}\,.$$

Note that even for this very small example with $n_{\boldsymbol{q}} = 2$ and only the consideration of the first five summands in the series expansion (3.11) this involves a quite large effort. To obtain one next coefficient in the series expansion a consistent initialization of a DAE system of index two is necessary which involves computation of higher order derivatives of all preceding coefficient functions.

In Figure 3.2 it is illustrated how the choice of initial values close to $\mathfrak{M}^\delta$ influences the growth in the norm of derivatives of $\boldsymbol{q}^\delta$ at $t = 0$. The singular force terms in (3.6) cause higher order derivatives of the the solution of the SSP system to grow by a factor of δ^{-1} with each time derivation. For $\delta = 10^{-4}$ which is a rather large value this already causes the seventh derivative of $\boldsymbol{q}^\delta(t)$ at $t = 0$ to be larger than 10^{20}. As a comparison we also added the norm of time derivatives of the *constrained* mechanical system indicated by "$\boldsymbol{q}$ (DAE)."

$\diamondsuit$

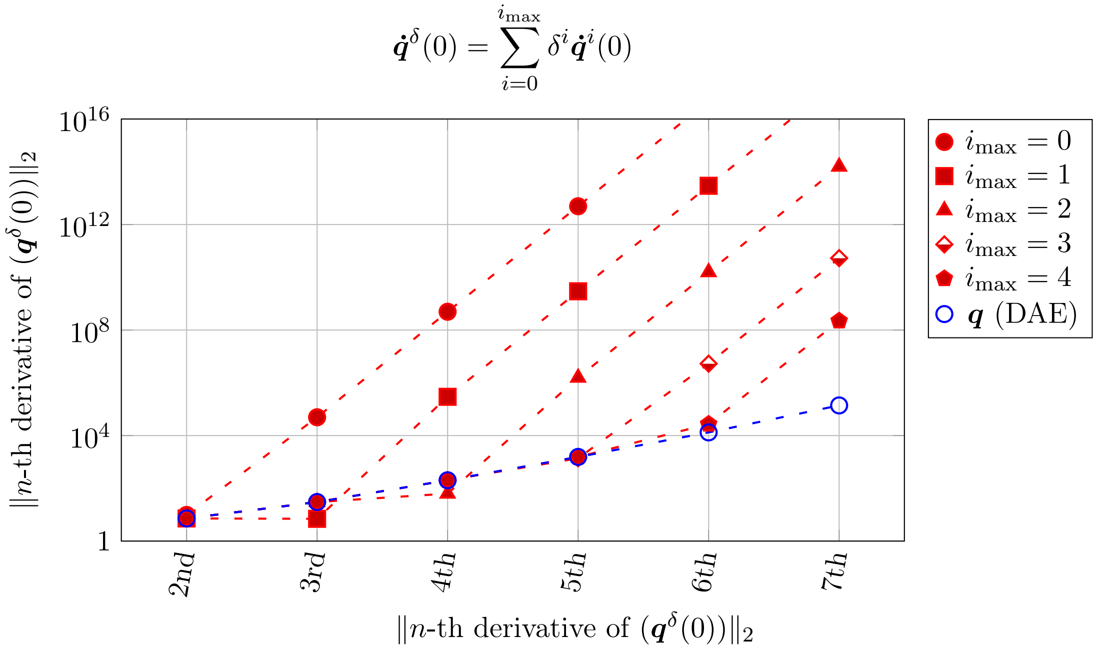
Remark 3.11 (Convergence results for Runge–Kutta methods for strongly damped mechanical systems)
Stumpp (2006) analyzes certain Runge–Kutta methods when applied to the problem class

$$\mathbf{M}(\boldsymbol{q}^\delta(t))\ddot{\boldsymbol{q}}^\delta(t) = \boldsymbol{f}(\boldsymbol{q}^\delta(t), \dot{\boldsymbol{q}}^\delta(t)) - \frac{1}{\delta}\mathbf{D}(\boldsymbol{q}(t))\dot{\boldsymbol{q}}^\delta(t)\,,$$

where $\mathbf{D}\colon \mathbb{R}^{n_{\boldsymbol{q}}} \to \mathbb{R}^{n_{\boldsymbol{q}} \times n_{\boldsymbol{q}}}$ maps onto symmetric positive semidefinite matrices of constant rank $n_{\boldsymbol{\lambda}}$. For $\mathbf{D}(\boldsymbol{q}) := [\mathbf{G}^\top\mathbf{G}](\boldsymbol{q})$ this exactly coincides with (3.6)*. If the Runge–Kutta method with invertible Runge–Kutta matrix $\mathbf{A}$ that has no eigenvalues on the negative real axis is of stage order $1 \le p_{\mathrm{s}} \le p-1$ for the order p of the method and the initial values $(\boldsymbol{q}_0^\delta, \boldsymbol{v}_0^\delta)^\top = (\boldsymbol{q}^\delta(t_0), \dot{\boldsymbol{q}}^\delta(t_0))^\top$ lie on $\mathfrak{M}^\delta$, it is shown that the errors when applied to* (3.6) *and to the index-2 DAE* (2.13) *are related like*

$$\boldsymbol{q}_n^\delta - \boldsymbol{q}^\delta(t_n) = \boldsymbol{q}_n - \boldsymbol{q}(t_n) + \mathcal{O}(\delta h^{p_{\mathrm{s}}})\,, \qquad \boldsymbol{v}_n^\delta - \dot{\boldsymbol{q}}^\delta(t_n) = \boldsymbol{v}_n - \dot{\boldsymbol{q}}(t_n) + \mathcal{O}(\delta h^{p_{\mathrm{s}}})\,, \tag{3.13}$$

for sufficiently small time step size $h > 0$. The initial values of (2.13) *are hereby defined by projecting $(\boldsymbol{q}_0^\delta, \boldsymbol{v}_0^\delta)^\top$ onto $\mathfrak{M}^{\mathrm{d}}$ using the projections $\boldsymbol{\pi}$ and $\mathbf{P}$ from Section 2.1.3, cf. Theorem 3.8. Moreover, this result can be extended to the case of initial values of* (3.6) *deviating from $\mathfrak{M}^{\mathrm{d}}$*



derivative	$i_{\max} = 0$	$i_{\max} = 1$	$i_{\max} = 2$	$i_{\max} = 3$	$i_{\max} = 4$	$\boldsymbol{q}$ (DAE)
2nd	9.8	7.2	7.2	7.2	7.2	7.2
3rd	49,367.2	7	30.2	30.3	30.3	30.2
4th	4.9e+8	2.9e+5	62.9	201.9	201.9	201.9
5th	4.9e+12	2.9e+9	1.6e+6	1,381	1,547.4	1,548
6th	4.9e+16	2.9e+13	1.6e+10	5.3e+6	27,225.7	13,197
7th	4.9e+20	2.9e+17	1.6e+14	5.3e+10	2.2e+8	1.4e+5

Figure 3.2: 'Smooth initial values' for the strongly damped pendulum example

by $\mathcal{O}(h)$. Using $\boldsymbol{\pi}$ and $\mathbf{P}$, it is possible to uniquely define initial values $(\boldsymbol{q}_0^{\delta,\mathrm{proj}}, \boldsymbol{v}_0^{\delta,\mathrm{proj}})^\top \in \mathfrak{M}^\delta$ and show the estimate

$$\|\boldsymbol{q}_n^\delta - \boldsymbol{q}_n^{\delta,\mathrm{proj}}\| + \|\boldsymbol{v}_n^\delta - \boldsymbol{v}_n^{\delta,\mathrm{proj}}\| \leq C(h\varrho^n + \delta^{p_\mathrm{s}+1}),$$

where $C > 0$ is a constant, $(\boldsymbol{q}_n^{\delta,\mathrm{proj}}, \boldsymbol{v}_n^{\delta,\mathrm{proj}})^\top$, $n = 1, 2, \ldots$, denote Runge–Kutta solutions with initialization $(\boldsymbol{q}_0^{\delta,\mathrm{proj}}, \boldsymbol{v}_0^{\delta,\mathrm{proj}})^\top$, and $|R(\infty)| < \varrho < 1$ is a constant that depends on the Runge–Kutta method and the relation h/δ. Practically, this imposes the existence of a constant $\tilde{C} > 0$ such that the above result is valid as long as $0 < \delta < \tilde{C}h$. For a comparison with the slow motion $(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t))^\top$ the above estimate leads to the result

$$\|\boldsymbol{q}_n^\delta - \boldsymbol{q}(t_n)\| + \|\boldsymbol{v}_n^\delta - \dot{\boldsymbol{q}}(t_n)\| \leq C(h\varrho^n + \delta + h^{p_\mathrm{DAE2}}),$$

where p_DAE2 denotes the order of the Runge–Kutta method when applied to (2.13)*. For more details see also the comprehensive discussion in (Stumpp, 2004).*

3.3 Stiff mechanical systems

In this section we draw our attention to the generalization of the spring pendulum example. As we have seen in Example 2.5, the physical principles are given by the Lagrange formalism as presented in Section 2.1.1 and it is not necessary to consider additional dissipative or external forces. To penalize deviations from the position constraint manifold $\mathfrak{M}^{\mathrm{s}}$, an additional potential term

$$\mathcal{U} := \frac{1}{2\varepsilon^2} \left\| \boldsymbol{g}(\boldsymbol{q}^{\varepsilon}(t)) \right\|_2^2 \tag{3.14}$$

is added to the Lagrangian $\mathcal{L}$ of the system. In the literature $\mathcal{U}$ is sometimes also called fictitious potential (Bayo et al., 1988, Arnold, 1989, Kurdila et al., 1993). Since $\mathcal{U}$ does not explicitly depend on the generalized velocities, from (2.1) and $\varepsilon^2 \cdot \frac{\partial \mathcal{U}}{\partial \boldsymbol{q}^{\varepsilon}} = \mathbf{G}(\boldsymbol{q}^{\varepsilon})\boldsymbol{g}(\boldsymbol{q}^{\varepsilon})$ we derive the equations of motion for stiff mechanical systems

$$\mathbf{M}(\boldsymbol{q}^{\varepsilon}(t))\ddot{\boldsymbol{q}}^{\varepsilon}(t) = \boldsymbol{f}(\boldsymbol{q}^{\varepsilon}(t), \dot{\boldsymbol{q}}^{\varepsilon}(t)) - \frac{1}{\varepsilon^2}\mathbf{G}^{\top}(\boldsymbol{q}^{\varepsilon}(t))\boldsymbol{g}(\boldsymbol{q}^{\varepsilon}(t)) \,. \tag{3.15}$$

Systems of this type 'naturally' appear in molecular dynamics where molecules form steep potential vaults almost constraining the motion and leading to high frequency oscillations. Another occurrence in molecular dynamics is bond stretching with bond angle bending where substitute models using constraints cannot be applied (Reich, 1995). As flexible multibody systems, the simulation of compressible fluids (Ebin, 1977) and, as our pendulum example, replacement and regularization models for constrained systems also fall into this framework, there has been much more research on stiff mechanical systems throughout the last decades than for strongly damped mechanical systems. Some researchers use the duality of (3.15) and the index-3 formulation of equations of motion (2.12) to study the numerical properties of DAE time integration methods; for HHT methods of the next chapter this has been carried out by Cardona and Géradin (1994) and for a large class of implicit Runge–Kutta methods by Lubich (1993), see Remark 3.20 below.

To classify (3.15) as a singular SPP, we use the partitioning $\boldsymbol{y}^{\varepsilon} := \boldsymbol{q}^{\varepsilon}$, $\boldsymbol{z}^{\varepsilon} := \boldsymbol{v}^{\varepsilon} = \dot{\boldsymbol{q}}^{\varepsilon}$ and obtain a system of the form (3.3) with ε being replaced by ε^2 and

$$\frac{\partial \boldsymbol{\psi}(\boldsymbol{y}^{\varepsilon}, \boldsymbol{z}^{\varepsilon})}{\partial \boldsymbol{z}^{\varepsilon}} = -\frac{\partial [\mathbf{M}^{-1}\mathbf{G}^{\top}\boldsymbol{g}](\boldsymbol{q}^{\varepsilon})}{\partial \boldsymbol{v}^{\varepsilon}} + \mathcal{O}(\varepsilon^2) = \mathcal{O}(\varepsilon^2) \,.$$

In the limit case $\varepsilon \to 0$, rank-deficiency, even vanishing, is observed. Thus, (3.15) is also a singular SPP. Equivalently to (3.7) we can again introduce an artificial Lagrange multiplier $\boldsymbol{\lambda}^{\varepsilon}(t) \in \mathbb{R}^{n_{\boldsymbol{\lambda}}}$ and obtain *Hairer's reformulation* (Hairer et al., 1989a, Lubich, 1993)

$$\mathbf{M}(\boldsymbol{q}^{\varepsilon}(t))\ddot{\boldsymbol{q}}^{\varepsilon}(t) = \boldsymbol{f}(\boldsymbol{q}^{\varepsilon}(t), \dot{\boldsymbol{q}}^{\varepsilon}(t)) - \mathbf{G}^{\top}(\boldsymbol{q}^{\varepsilon}(t))\boldsymbol{\lambda}^{\varepsilon}(t) \,, \tag{3.16a}$$

$$\varepsilon^2 \boldsymbol{\lambda}^{\varepsilon}(t) = \boldsymbol{g}(\boldsymbol{q}^{\varepsilon}(t)) \,, \tag{3.16b}$$

which can easily be seen to be, for each finite value of $\varepsilon > 0$, a DAE of index one, since a differentiation of (3.16b) leads to the differential equation

$$\dot{\boldsymbol{\lambda}}^{\varepsilon}(t) = \frac{1}{\varepsilon^2}\mathbf{G}(\boldsymbol{q}^{\varepsilon}(t))\dot{\boldsymbol{q}}^{\varepsilon}(t)$$

for determining $\boldsymbol{\lambda}^{\varepsilon}(t)$. In view of the index-1 condition (2.8) above, the matrix $\boldsymbol{\psi}_{\boldsymbol{z}} = -\varepsilon^2\mathbf{I}$ is evidently invertible. It is, nevertheless, a singularly perturbed (DAE) problem and the main advantage of this reformulation lies, from the viewpoint of numerical analysis, in the closer connection to the standard form of the index-3 DAE (2.12) and computational advantages. In conclusion, (3.16) is a singular SPP as (3.7) and we will also refer to it as the 'index-1 formulation' of the SPP (3.15).

Remark 3.12
It might at first glance seem rather arbitrary to scale the penalizing potential by a factor of ε^{-2} *instead of just* ε^{-1}*, corresponding to the factor* δ^{-1} *in* (3.6)*. It is nevertheless beneficial to hold*

to this approach for more than just consistency with the literature: When the stiff forces are scaled by ε^{-2}, the oscillations induced by the leading linear part scale with ε rather than $\sqrt{\varepsilon}$. As a result, the derivatives of $\boldsymbol{q}^{\varepsilon}(t)$, cf. Example 3.10(b), grow with each time derivation by a factor of ε^{-1} as the δ^{-1} in Figure 3.2. Even more important, the singular behavior imposes a restriction on the ratio of time step size h for numerical methods and the penalty parameter: In Remark 3.11 we saw that the results are valid for $0 < \delta < \tilde{C}h$. The corresponding results for stiff mechanical systems are due to the work of Lubich (1993) and will be summarized in Remark 3.20. If one is to resolve the oscillations, we have already mentioned in Section 2.2.3 that the time step size needs to be of the same order as the frequency of the oscillations. Either way, a scaling with ε^{-2} appears to be the natural choice.

Nevertheless when judging the numerical results later on, one should always keep in mind that a reduction of the penalty parameter ε for the stiff systems by a factor $c \in \mathbb{R}$ corresponds to a multiplication of the perturbation parameter δ by $\sqrt{c}$ – judging from the absolute value of the penalizing force terms.

Remark 3.13 ('Takens-chaos': Analytically inherent singular behavior)
Proving the physically intuitive fact that for conservative systems and initial values that are consistent with the corresponding constrained mechanical systems to (3.15) *approach solutions of* (2.12) *for $\varepsilon \to 0$ has long been an open mathematical question. Rubin and Ungar (1957) gave a first mathematically rigorous proof using functional analytic techniques as the Arzéla/Ascoli Theorem for convergence in function spaces. They already pointed out the importance of the analysis in case of consistent initial values as opposed to those violating velocity constraints, as then only weak or no convergence can be proven.*

There have been many extensions and alternative proofs employing methods from various fields as averaging (Arnold, 1989) or energy principles (Kurdila et al., 1993). To our knowledge the first proof based on singular perturbation theory and involving the existence of a series expansion like in (3.11) *has been published by Lötstedt (1979). This approach bears the advantages that not only it includes the nontrivial observation that solutions starting in $T\,\mathfrak{M}^{\mathrm{s}}$ are not optimally smooth in the sense that of all possible initial configurations they are the least oscillating, but also provides techniques to construct such initial values. In Theorem 3.16 we will use the formulation and proof of the Rubin–Ungar Theorem as presented by Lubich (1993) which is in parts based on the work of Lötstedt (1979) before we construct smooth(er) initial values for the stiff spring pendulum in Example 3.17.*

In contrast to the analytic transformation from Remark 3.6, in case of stiff mechanical systems it is not known in general how to find a smooth coordinate change such that (3.15) *may be interpreted as a regular SPP. On the one hand, this is already evident as the spectrum of stiff mechanical systems for very small perturbation parameter ε lies very close to the imaginary axis and so does not impose stable, i. e., attractive, behavior of the dynamical system. Moreover, in case of initial values deviating too much from the constraints (but still bounded energy in the system), there are examples (Bornemann, 1998, Chap. 2 §4) where the limiting behavior is no longer uniquely determined since resonances lead to chaotic oscillations. This phenomenon was first studied in detail by Takens (1980) and is therefore named 'Takens-chaos'.*

As in Lemma 3.7 it is also possible to reformulate the equations of motion of stiff mechanical systems such that the singular forces enter linearly and define a partitioning into stiff and nonstiff variables. Since $\mathcal{U}$ only depends on position variables, the position-velocity relation remains unchanged.

Lemma 3.14 (Alternative formulation of stiff mechanical systems (Lubich, 1993, Lemma 2.1))
For given $\boldsymbol{q}^0 \in \mathfrak{M}^{\mathrm{s}}$, there exists, locally but independent of ε, a coordinate change $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{q})$, or $\boldsymbol{q} = \boldsymbol{q}(\boldsymbol{z})$, respectively, with $\boldsymbol{q}(\boldsymbol{0}) = \boldsymbol{q}^0$ and as often continuously differentiable as $\nabla^2\mathcal{U}$, such

that in the new coordinates $\boldsymbol{z}$ the potential takes the form

$$\mathcal{U}(\boldsymbol{q}(\boldsymbol{z})) = \frac{1}{2\varepsilon^2}\|\boldsymbol{z}^{\perp}\|_2^2\,,$$

where $\boldsymbol{z} = (\boldsymbol{z}^{\|}, \boldsymbol{z}^{\perp})^{\top}$ is partitioned in components $\boldsymbol{z}^{\|} \in \mathbb{R}^{n_{\boldsymbol{q}}-n_{\boldsymbol{\lambda}}}$ parallel to $\mathfrak{M}^{\mathrm{s}}$ and $\boldsymbol{z}^{\perp} \in \mathbb{R}^{n_{\boldsymbol{\lambda}}}$ orthogonal to it. In terms of $\boldsymbol{z}(t)$, the equations of motion of the stiff mechanical system read

$$\hat{\mathbf{M}}(\boldsymbol{z}(t))\ddot{\boldsymbol{z}}(t) = \hat{\boldsymbol{f}}(\boldsymbol{z}(t), \dot{\boldsymbol{z}}(t)) - \frac{1}{\varepsilon^2}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_{\boldsymbol{\lambda}}}\end{pmatrix}\boldsymbol{z}(t) \tag{3.17}$$

with $\hat{\mathbf{M}}(\boldsymbol{z}) := (\partial\boldsymbol{q}/\partial\boldsymbol{z})^{\top}\mathbf{M}(\boldsymbol{q}(\boldsymbol{z}))(\partial\boldsymbol{q}/\partial\boldsymbol{z})$, $\hat{\boldsymbol{f}}(\boldsymbol{z}, \dot{\boldsymbol{z}}) := (\partial\boldsymbol{q}/\partial\boldsymbol{z})\boldsymbol{f}(\boldsymbol{q}(\boldsymbol{z}), \dot{\boldsymbol{q}}(\boldsymbol{z}, \dot{\boldsymbol{z}}))$.

Remark 3.15 (Linear singular perturbations and flexible multibody systems)
For linear elasticity problems in flexible multibody dynamics, see Example 3.19 below, the structure of (3.17) *is sometimes already the situation at hand, if the elastic deformations are defined such that their vanishing implies that the system moves like the gross motion alone, i. e., like a rigid system without flexible components (Simeon, 2013).*

This problem class is also appealing because the Hessian of $\mathcal{U}$ is constant. We will see in the following theorem that, as for the strongly damped counterpart in Theorem 3.8, consistent initial values (for the DAE case) do not necessarily result in a smooth solution. In molecular and flexible multibody dynamics this matter can sometimes be diminished by adding so-called correcting potentials *(Reich, 1995). For problems with constant Hessian $\nabla^2\mathcal{U}$ those correction terms vanish, i. e., the solution is already smooth.*

Theorem 3.16 (Smooth motion (Lubich, 1993))
For each pair $(\boldsymbol{q}_0^0, \dot{\boldsymbol{q}}_0^0)^{\top}$ satisfying

$$\boldsymbol{g}(\boldsymbol{q}_0^0) = \mathbf{G}(\boldsymbol{q}_0^0)\dot{\boldsymbol{q}}_0^0 = \mathbf{0}$$

and arbitrarily given $i_{\max} > 0$ there exist $(\boldsymbol{q}_0^{\varepsilon}, \dot{\boldsymbol{q}}_0^{\varepsilon})^{\top}$ which are unique to $\mathcal{O}(\varepsilon^{2\cdot i_{\max}})$, and with

$$\boldsymbol{q}_0^0 - \boldsymbol{q}_0^{\varepsilon},\ \dot{\boldsymbol{q}}_0^0 - \dot{\boldsymbol{q}}_0^{\varepsilon} \in \mathcal{O}(\varepsilon^2) \cap \left(T_{\boldsymbol{q}_0^0}\,\mathfrak{M}^{\mathrm{s}}\right)^{\mathbf{M}(\boldsymbol{q}_0^0)\text{-}\perp} \tag{3.18}$$

such that the solution $(\boldsymbol{q}^{\varepsilon}(t), \dot{\boldsymbol{q}}^{\varepsilon}(t))^{\top}$ of (3.15) with initial values $(\boldsymbol{q}_0^{\varepsilon}, \dot{\boldsymbol{q}}_0^{\varepsilon})^{\top}$ is smooth and allows for an expansion of the form

$$\begin{aligned}\boldsymbol{q}^{\varepsilon}(t) &= \boldsymbol{q}^0(t) + \varepsilon^2\boldsymbol{q}^1(t) + \ldots + \varepsilon^{2\cdot i_{\max}}\boldsymbol{q}^{i_{\max}}(t) + \mathcal{O}(\varepsilon^{2\cdot i_{\max}+2})\,,\\ \dot{\boldsymbol{q}}^{\varepsilon}(t) &= \dot{\boldsymbol{q}}^0(t) + \varepsilon^2\dot{\boldsymbol{q}}^1(t) + \ldots + \varepsilon^{2\cdot i_{\max}}\dot{\boldsymbol{q}}^{i_{\max}}(t) + \mathcal{O}(\varepsilon^{2\cdot i_{\max}+2})\,,\end{aligned} \tag{3.19}$$

where $\boldsymbol{q}^i(t)$, $i \geq 0$, are ε-independent and exist on finite time intervals. In particular, $\boldsymbol{q}^0(t) = \boldsymbol{q}(t)$, $t \in [t_0, t_{\mathrm{end}}]$, is the solution of the index-3 equations of motion of the corresponding constrained system (2.12).

All pairs $(\boldsymbol{q}_0^{\varepsilon}, \dot{\boldsymbol{q}}_0^{\varepsilon})^{\top}$ form a $2(n_{\boldsymbol{q}} - n_{\boldsymbol{\lambda}})$-dimensional manifold $\mathfrak{M}^{\varepsilon} \subset \mathbb{R}^{2n_{\boldsymbol{q}}}$, which is invariant under the flow of (3.15) up to terms of order $\mathcal{O}(\varepsilon^{2\cdot i_{\max}+2})$.

Instead of giving a formal proof to this theorem, we come back once again to the planar stiff pendulum. In the following example we show that there is no need to compute an analytic expression for the solution to get initial values that allow for a smooth analytic solution, i. e., one with arbitrarily many, ε-independently-bounded derivatives. It is, on the contrary, possible to derive such initial values using only the involved functions $\mathbf{M}$, $\boldsymbol{f}$ and $\boldsymbol{g}$ and their derivatives, respectively. One should nevertheless keep in mind that for a large scale computer model in a technical simulation it is prone trying to find initial values that lie ‘exactly on’ $\mathfrak{M}^{\varepsilon}$, yet alone sufficiently close to it such that the numerical procedures are not negatively influenced.

Example 3.17 (Smooth motion of the stiff spring pendulum)
A formal insertion of the series ansatz (3.19) into the equations of motion for the stiff mechanical system and expansion with respect to powers of ε^2 allows to equate series coefficients recursively: The ε^{-2}-term vanishes iff

$$\mathbf{G}^\top(\boldsymbol{q}^0)\boldsymbol{g}(\boldsymbol{q}^0) = \mathbf{0} \quad \Leftrightarrow \quad \boldsymbol{g}(\boldsymbol{q}^0) = \mathbf{0}\,.$$

Equating the ε^0-terms

$$\mathbf{M}(\boldsymbol{q}^0)\ddot{\boldsymbol{q}}^0 = \boldsymbol{f}(\boldsymbol{q}^0, \dot{\boldsymbol{q}}^0) - \mathbf{G}^\top(\boldsymbol{q}^0)\mathbf{G}(\boldsymbol{q}^0)\boldsymbol{q}^1\,,$$

we almost obtain the descriptor form already. To attain a uniquely solvable system we formally introduce the variable $\mathbb{R}^{n_{\boldsymbol{\lambda}}} \ni \boldsymbol{\lambda}^0(t) := \mathbf{G}(\boldsymbol{q}^0(t))\boldsymbol{q}^1(t)$ and arrive at (2.12) for $\boldsymbol{q} = \boldsymbol{q}^0(t)$. Considering the solution $(\boldsymbol{q}^0(t), \dot{\boldsymbol{q}}^0(t), \ddot{\boldsymbol{q}}^0(t))^\top$ to be known, equating the ε^2-terms gives another system of the same structure

$$\mathbf{M}(\boldsymbol{q}^0) \cdot \ddot{\boldsymbol{q}}^1 = \boldsymbol{f}^1(\boldsymbol{q}^1, \dot{\boldsymbol{q}}^1; \boldsymbol{q}^0, \dot{\boldsymbol{q}}^0, \ddot{\boldsymbol{q}}^0) - \mathbf{G}^\top(\boldsymbol{q}^0)\underbrace{\mathbf{G}(\boldsymbol{q}^0)\boldsymbol{q}^2}_{=:\boldsymbol{\lambda}^1}, \tag{3.20}$$

where $\boldsymbol{f}^1$ is linear in $\boldsymbol{q}^1$, $\dot{\boldsymbol{q}}^1$. Together with the above definition of $\boldsymbol{\lambda}^0$ as a constraint equation this is again an index-3 differential-algebraic system providing a unique solution $\boldsymbol{q}^1$. By induction, this procedure may be extended to arbitrary order $\boldsymbol{q}^k$. Note however, that for the definition of $\boldsymbol{q}^2$ (or—more precisely—the definition of $\boldsymbol{f}^2$), the fourth order derivatives of $\boldsymbol{q}^0$ and the second derivatives of $\boldsymbol{\lambda}^0$ are needed. Altogether, the definition of $\boldsymbol{q}^k$ involves the solution of a sequence of $k-1$ index-3 or one (large) index-$(2k+3)$ system and is therefore beyond the capacity of what one can expect from a numerical algorithm. After the formal derivation of $\boldsymbol{q}^k$, $k = 0, 1, \ldots$, finishing a formal proof of Theorem 3.16 mostly copes with the estimation of the remainder in truncated series expansions using Lemma 3.14 above to basically deal with linear problems. The entire proof can be found in (Lubich, 1993, Theorem 2.2).

Coming back to the pendulum example, for $\boldsymbol{q}(0) = \frac{1}{2}(\sqrt{2}, \sqrt{2})^\top$, $\dot{\boldsymbol{q}}(0) = (-1, 1)^\top$, we get (1.1a), (1.1b) to determine $\boldsymbol{q}^0$. The solution of the saddle-point problem (2.15) can be used to obtain

$$\begin{aligned}
\ddot{\boldsymbol{q}}^0(0) &= \begin{pmatrix} \frac{g_{\text{grav}}}{2} - \sqrt{2} \\ -\frac{g_{\text{grav}}}{2} - \sqrt{2} \end{pmatrix}, \\
\lambda^0(0) &= \left.\frac{(q_y^0)^2\left((\dot{q}_x^0)^2 - g_{\text{grav}} q_y^0\right) + (q_x^0)^2\left(\dot{(q_y^0)}^2 - g_{\text{grav}} q_y^0\right) - 2q_x^0 q_y^0 \dot{q}_x^0 \dot{q}_y^0}{\left((q_x^0)^2 + (q_y^0)^2\right)^{3/2}}\right|_{t=0} = 2 - \tfrac{\sqrt{2}}{2} g_{\text{grav}}\,.
\end{aligned}$$

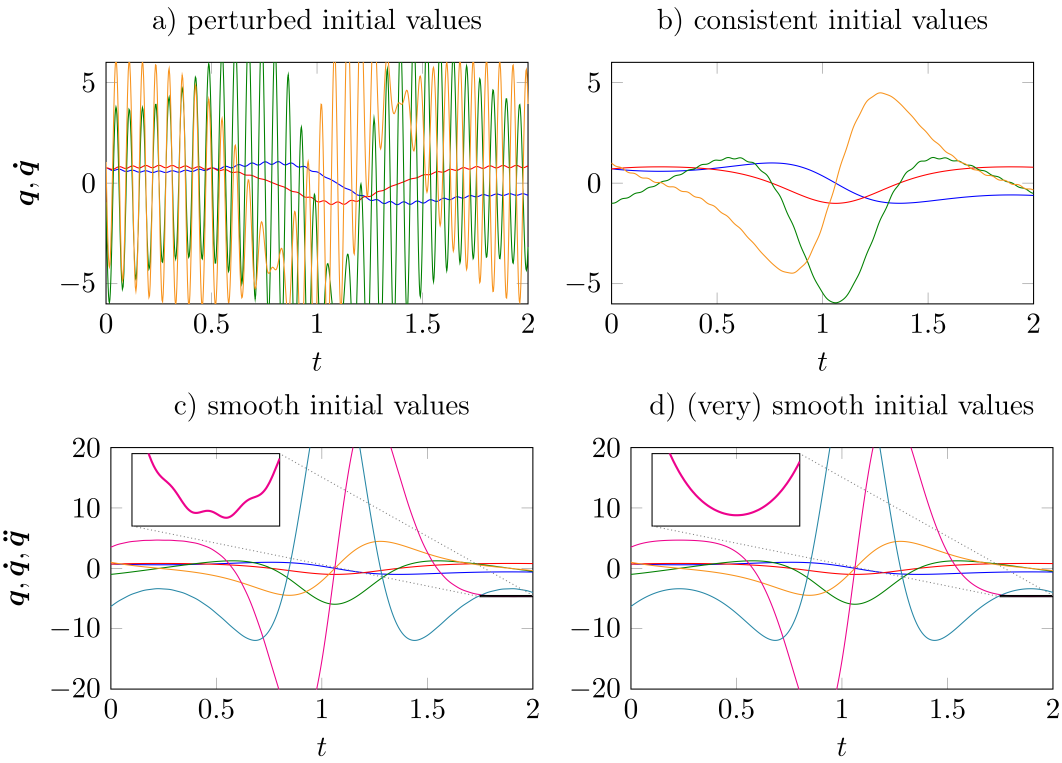
To derive initial values for $\boldsymbol{q}^1$, we use (3.20) which for this very simple problem already reads

$$\begin{pmatrix} \ddot{q}_x^1 \\ \ddot{q}_y^1 \end{pmatrix} = \frac{\lambda^0}{\left((q_x^0)^2 + (q_y^0)^2\right)^{3/2}} \begin{pmatrix} q_y^0(q_y^0 q_x^1 - q_x^0 q_1^1) \\ q_x^0(q_x^0 q_y^1 - q_y^0 q_x^1) \end{pmatrix} - \mathbf{G}^\top(\boldsymbol{q}^0)\lambda^1$$

and an index reduction to obtain initial values for the system with $\mathbf{G}(\boldsymbol{q}^0)\boldsymbol{q}^1 = \lambda^0$ includes two time differentiations of $\lambda^0(t)$ and so on. Carefully equating terms with ε^4 and always adding a condition that the initial values lie in the $\mathbf{M}(\boldsymbol{q})$-orthogonal complement of $\mathfrak{M}^{\text{s}}$ gives the following set of initial values:

$$\begin{aligned}
\boldsymbol{q}^\varepsilon(0) &= \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix} + \varepsilon^2 \begin{pmatrix} \sqrt{2} - \frac{g_{\text{grav}}}{2} \\ \sqrt{2} - \frac{g_{\text{grav}}}{2} \end{pmatrix} + \varepsilon^4 \left(-\frac{3g_{\text{grav}}}{\sqrt{2}} - \frac{3g_{\text{grav}}}{\sqrt{2}}\right) + \mathcal{O}(\varepsilon^6)\,, \\
\dot{\boldsymbol{q}}^\varepsilon(0) &= \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \varepsilon^2 \begin{pmatrix} -\frac{3g_{\text{grav}}^2 + 4\sqrt{2}g_{\text{grav}} + 8}{2\sqrt{2}} \\ -\frac{3g_{\text{grav}}^2 + 4\sqrt{2}g_{\text{grav}} + 8}{2\sqrt{2}} \end{pmatrix} + \varepsilon^4 \begin{pmatrix} \frac{9}{2} g_{\text{grav}}\left(g_{\text{grav}} + 2\sqrt{2}\right) \\ \frac{9}{2} g_{\text{grav}}\left(g_{\text{grav}} + 2\sqrt{2}\right) \end{pmatrix} + \mathcal{O}(\varepsilon^6)\,.
\end{aligned} \tag{3.21}$$

Note that the orthogonality conditions in (3.18) imply the unique definition of corresponding smooth initial values and coincide with $(\boldsymbol{q}_0^0, \dot{\boldsymbol{q}}_0^0)^\top$ being the projected values of $(\boldsymbol{q}_0^\varepsilon, \dot{\boldsymbol{q}}_0^\varepsilon)^\top$ using the techniques of Section 2.1.3. Notice also that the same result can be acquired using Hairer's reformulation (3.16) and a series ansatz for $\boldsymbol{\lambda}^\varepsilon$. This approach is arguably easier to use from a practical point of view because the definition of $\boldsymbol{\lambda}^k$ does not have to be imposed in each induction step.


a) perturbed initial values

b) consistent initial values

c) smooth initial values

d) (very) smooth initial values

Figure 3.3: Numerical solution of pendulum equations using `ode113.m` and different initial values

In Figure 3.3 we illustrate the influence of the choice of initial values on the behavior of the solution. We used the Matlab$^{\text{TM}}$ integration method `ode113.m` (Shampine and Reichelt, 1997) with very low tolerances `'AbsTol'='RelTol'=1e-12`, `'InitialStep'=1e-14` and the moderate value $\varepsilon = 10^{-2}$ to compute very accurate solutions to the stiff system with (a) initial values that were perturbed by $0.05 \cdot (1,1)^\top$ on position and velocity level, respectively, (b) consistent initial values, (c) using the initial values from (3.21) truncating after the ε^2-terms and (d) with the computed smooth initial values. The number of function evaluations of `ode113.m`—the probably most viable way to measure computational cost–dropped from 4924 for perturbed values, to 3339 for consistent initialization to 2331 and 1653 respectively for the smooth initial values. To obtain the results in the lower row of Figure 3.3, we enlarged the system by adding new variables for the accelerations $\ddot{\boldsymbol{q}}^\varepsilon$. The numerical cost was comparable to the above experiment (5993, 4527, 3315, 2009) but zooming in for $t \in [1.75, 2]$, $\ddot{q}_x^\varepsilon \in [-4.7, -4.5]$ reveals the stronger tendency towards oscillations for the third choice. One should nevertheless emphasize that Theorem 3.16 is mainly of mathematical interest and that for practical computations it is neither possible to obtain smooth initial values nor to hope that the computation is accurate enough to validate the assertion. $\diamondsuit$

Corollary 3.18 (Rubin–Ungar Theorem (II))
The *well-defined* solutions of the equations of motion of stiff mechanical systems (3.15) with consistent initial values $\boldsymbol{q}(t_0)$, $\dot{\boldsymbol{q}}(t_0)$ approach for $\varepsilon \to 0$ uniformly on the finite time interval $[t_0, t_{\text{end}}]$ the solutions of the equations of motion in descriptor form (2.12). The difference of approximates and limit solutions $\|\boldsymbol{q}^\varepsilon(t) - \boldsymbol{q}(t)\|$ remain in $\mathcal{O}(\varepsilon^2)$, $\|\dot{\boldsymbol{q}}^\varepsilon(t) - \dot{\boldsymbol{q}}(t)\|$ in $\mathcal{O}(\varepsilon)$. The result remains true if the deviation of $\boldsymbol{q}_0^\varepsilon$ and $\dot{\boldsymbol{q}}_0^\varepsilon$ from $\mathfrak{M}^{\text{s}}$ and $T_{\boldsymbol{\pi}\boldsymbol{q}_0^\varepsilon}\mathfrak{M}^{\text{s}}$ is $\mathcal{O}(\varepsilon^2)$, $\mathcal{O}(\varepsilon)$ respectively.

Note that even for consistent initial values, Theorem 3.16 does not provide a $\mathcal{O}(\varepsilon^2)$-estimate for the errors $\|\dot{\boldsymbol{q}}^\varepsilon(t_n) - \dot{\boldsymbol{q}}(t_n)\|$ on velocity level. This is only true for smooth solutions. In particular, this implies that

$$\|\mathbf{P}(\boldsymbol{q}(t))\dot{\boldsymbol{q}}^\varepsilon(t) - \dot{\boldsymbol{q}}(t)\| = \|\mathbf{P}(\boldsymbol{q}(t))(\dot{\boldsymbol{q}}^\varepsilon(t) - \dot{\boldsymbol{q}}(t))\| = \mathcal{O}(\varepsilon^2), \text{ while } \|\dot{\boldsymbol{q}}^\varepsilon(t) - \dot{\boldsymbol{q}}(t)\| = \mathcal{O}(\varepsilon) \text{ in general.} \tag{3.22}$$

A more detailed analysis is given by Bornemann (1998).

Example 3.19 (Stiff mechanical systems in flexible multibody systems)
In the simulation of (bio-) mechanical systems the consideration of just rigid bodies is often insufficient to determine the time evolution of a complex system. The simulation of these *flexible multibody systems* often leads to stiff mechanical systems which is why we give a very concise overview here. Flexible structures or tissue can be described using the *Navier–Lamé equations*

$$\begin{aligned} \rho\ddot{\boldsymbol{u}}(\boldsymbol{x},t) &= \mathbf{div}\boldsymbol{\Sigma}(\boldsymbol{u}(\boldsymbol{x},t)) + \boldsymbol{\beta}(\boldsymbol{x},t)\,,\ \boldsymbol{x} \in \Omega\,,\ t \in [t_0, t_{\text{end}}]\,, \\ \boldsymbol{u}(\boldsymbol{x},t) &= \boldsymbol{u}_0(\boldsymbol{x},t)\,,\ \boldsymbol{x} \in \Gamma^0\,, \quad \boldsymbol{\Sigma}(\boldsymbol{u}(\boldsymbol{x},t))\boldsymbol{n}(\boldsymbol{x}) = \boldsymbol{\tau}(\boldsymbol{x},t)\,,\ \boldsymbol{x} \in \Gamma^1\,, \end{aligned} \tag{3.23}$$

of structural dynamics or variants of it. In PDE (3.23), $\boldsymbol{u}\colon\ \mathbb{R}^3 \times [t_0, t_{\text{end}}] \to \mathbb{R}^3$ is the *displacement* of each point of the body, ρ is the mass density and $\boldsymbol{\Sigma} := \Lambda_1 \cdot (\text{trace}(\boldsymbol{\epsilon}))\mathbf{I} + 2\Lambda_2 \cdot \boldsymbol{\epsilon}\colon\ \mathbb{R}^3 \to \mathbb{R}^{3\times3}$ is the *St. Venant–Kirchhoff stress tensor* with *Lamé constants* Λ_1, $\Lambda_2 \in \mathbb{R}$ and the *Green–Lagrangian strain tensor* $\boldsymbol{\epsilon} := \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^\top + \nabla\boldsymbol{u}^\top\nabla\boldsymbol{u})$. Γ^0 and Γ^1 with $\Gamma^0 \cup \Gamma^1 = \partial\Omega$ are the Dirichlet and Neumann boundaries of the bounded domain $\Omega \subset \mathbb{R}^3$ that represents the flexible body and $\boldsymbol{\beta}\colon\ \mathbb{R}^3 \times [t_0, t_{\text{end}}] \to \mathbb{R}^3$ is the vector of external and internal forces. Through $\boldsymbol{\beta}$, the Dirichlet data $\boldsymbol{u}_0$, and the Neumann velocity field $\boldsymbol{\tau}$ the body is usually coupled to other structures in the ODE- or DAE-model. Depending on the geometry and coupling conditions there are many numerical approaches to (3.23). Most common is a semidiscretization technique (method-of-lines) where the PDE is transformed to a system of ODEs: Based on the *weak formulation*

$$\forall \boldsymbol{v} \in V_0 :\ \int_\Omega \rho\boldsymbol{v}^\top\ddot{\boldsymbol{u}}\ \mathrm{d}\boldsymbol{x} + \int_\Omega \boldsymbol{\Sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{v})\ \mathrm{d}\boldsymbol{x} = \int_\Omega \boldsymbol{v}^\top\boldsymbol{\beta}\ \mathrm{d}\boldsymbol{x} + \int_{\Gamma^1} \boldsymbol{v}^\top\boldsymbol{\tau}\ \mathrm{d}s\,, \tag{3.24}$$

or more abstract $\forall \boldsymbol{v} \in V_0 : \langle\rho\ddot{\boldsymbol{u}}, \boldsymbol{v}\rangle + a(\boldsymbol{u}, \boldsymbol{v}) = \langle\boldsymbol{l}, \boldsymbol{v}\rangle$ for a suitable Sobolev space V_0, nonlinear functionals $a(\cdot,\cdot)$ and $\langle\cdot,\cdot\rangle$, and an element $\boldsymbol{l} \in V_0$ dependent on $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, see (Simeon, 2013) for details.

The basis of the *Galerkin-approach* (Hughes, 1987) lies at restricting (3.24) to a finite dimensional ansatz space $V_{h,0}$ of appropriately chosen functions $\boldsymbol{v}_i$, $i = 1, \ldots, n_{\text{d}} := \dim V_{h,0}$, i. e.,

$$\forall \boldsymbol{v}_i \in V_{h,0} :\ \langle\rho\ddot{\boldsymbol{u}}, \boldsymbol{v}_i\rangle + a(\boldsymbol{u}, \boldsymbol{v}_i) = \langle\boldsymbol{l}, \boldsymbol{v}_i\rangle\,.$$

The displacement $\boldsymbol{u}$ is then approximated by a linear combination $\boldsymbol{u}_h(t,\boldsymbol{x}) := \sum_{j=1}^{n_d} q_j(t) \cdot \boldsymbol{v}_j(\boldsymbol{x})$ of the elements of $V_{h,0}$. If internal damping is neglected, this leads to a second order system

$$\mathbf{M}(\boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) + \mathbf{K}(\boldsymbol{q}(t))\boldsymbol{q}(t) = \boldsymbol{f}(t, \boldsymbol{q}(t))\,,\ t \in [t_0, t_{\text{end}}]\,, \tag{3.25}$$

where $\boldsymbol{q}(t) = (q_i)_{i=1,\dots,n_\mathrm{d}}$ contains the coefficients in the approximation of $\boldsymbol{u}$ by $\boldsymbol{u}_h$ and

$$\mathbf{M} = \left(\langle \rho \boldsymbol{v}_j, \boldsymbol{v}_i \rangle\right)_{i,j=1,\dots,n_d}, \quad \mathbf{K} = \left(a(\boldsymbol{v}_j, \boldsymbol{v}_i)\right)_{i,j=1,\dots,n_d}, \quad \boldsymbol{f} = \left(\langle \boldsymbol{l}, \boldsymbol{v}_i \rangle\right)_{i=1,\dots,n_d}$$

are the locally assembled *mass* and *stiffness matrix* and $\boldsymbol{f}$ is called *load vector*. In many applications the gross motion is much larger than the deformations $\boldsymbol{u}$ and it suffices to consider linear elasticity: The strain tensor is truncated after the second summand and $a(\cdot,\cdot)$ and $\langle\cdot,\cdot\rangle$ become bilinear forms. As a result, $\mathbf{M}$ and $\mathbf{K}$, sometimes even $\boldsymbol{f}$, are independent of the state vector $\boldsymbol{q}$ and need to be assembled only once.

Whether or not (3.25) may be regarded as a stiff mechanical system in the above sense may now depend on various components: The decisive factor is whether certain time scales in the system are considerably separated from others, or, for linear elasticity, there is a substantial gap in the eigenvalue spectrum of $(\mathbf{M}, \mathbf{K})$. In practice, this depends on model parameters as material or geometry and the ansatz functions/the semidiscretization procedure (including the triangulation of Ω). For rod or shell models the separation is typically given since bending modes usually show much smaller frequencies than shearing or elongation modes. Becker (2012) studies almost incompressible media where the bulk modulus is very large and the system may therefore be treated as a singularly perturbed system. For an application in biological tissue simulation we refer to Simeon et al. (2009).

At last, we note that linear elasticity theory leads to linear stiff potential forces and corresponds to a quadratic penalizing potential which bears several advantages from a computational viewpoint. Note also that the discretization in space introduces an additional spatial error in the numerical solution. ♢

Remark 3.20 (Convergence results for Runge–Kutta methods and stiff mechanical systems) *In (Lubich, 1993) a large class of Runge–Kutta methods is analyzed for stiff mechanical systems of the form*

$$\mathbf{M}(\boldsymbol{q}^\varepsilon(t))\ddot{\boldsymbol{q}}^\varepsilon(t) = \boldsymbol{f}(\boldsymbol{q}^\varepsilon(t), \dot{\boldsymbol{q}}^\varepsilon(t)) - \frac{1}{\varepsilon^2}\nabla\,\mathcal{U}(\boldsymbol{q}^\varepsilon(t))\,,$$

where the (slightly more general) potential $\mathcal{U}$ attains a local minimum along an $(n_{\boldsymbol{q}} - n_{\boldsymbol{\lambda}})$-dimensional manifold. The results we present in the next chapters for Newmark integration methods also apply in this setting.

Let a Runge–Kutta method with classical order p and stage order $1 \le p_\mathrm{s} \le p$ be given that is I-stable and has an invertible Runge–Kutta matrix $\mathbf{A}$ with no eigenvalues on the imaginary axis. If the initial values $(\boldsymbol{q}_0^\varepsilon, \boldsymbol{v}_0^\varepsilon)^\top$ lie on the manifold $\mathfrak{M}^\varepsilon$ of smooth motion we get the equivalent to the estimate (3.13)

$$\boldsymbol{q}_n^\varepsilon - \boldsymbol{q}^\varepsilon(t_n) = \boldsymbol{q}_n^0 - \boldsymbol{q}(t_n) + \mathcal{O}(\varepsilon^2 h^{p_\mathrm{s}-2})\,, \qquad \boldsymbol{v}_n^\varepsilon - \dot{\boldsymbol{q}}^\varepsilon(t_n) = \boldsymbol{v}_n^0 - \dot{\boldsymbol{q}}(t_n) + \mathcal{O}(\varepsilon^2 h^{p_\mathrm{s}-2})\,,$$

to relate errors of the method for application to (3.15) *and those for* (2.12) *with initial values attained by mass orthogonal projection. When the initial values deviate from $\mathfrak{M}^\varepsilon$ by terms of magnitude $\mathcal{O}(h^2)$ for position and $\mathcal{O}(h)$ for velocity coordinates, the discrete dynamical system defined by the Runge–Kutta method has an attractive invariant manifold and one can show the estimates*

$$\|\boldsymbol{q}_n^\varepsilon - \boldsymbol{q}_n^{\varepsilon,\mathrm{proj}}\| + \|\boldsymbol{v}_n^\varepsilon - \boldsymbol{v}_n^{\varepsilon,\mathrm{proj}}\| \le \begin{cases} C(h\varrho^n + \varepsilon^{p_\mathrm{s}}) & \textit{for even } p_\mathrm{s}\,, \\ C(h\varrho^n + h\varepsilon^{p_\mathrm{s}-1}) & \textit{for odd } p_\mathrm{s}\,, \end{cases} \tag{3.26}$$

where $(\boldsymbol{q}_0^{\varepsilon,\mathrm{proj}}, \boldsymbol{v}_0^{\varepsilon,\mathrm{proj}})^\top \in \mathfrak{M}^\varepsilon$ are uniquely defined by the projection to consistent values $\boldsymbol{\pi}(\boldsymbol{q}_0^\varepsilon)$, $\mathbf{P}\boldsymbol{v}_0^\varepsilon$ and their counterparts on $\mathfrak{M}^\varepsilon$. $\varrho < 1$ is again a number that depends on the method and the ratio of time step size h and penalty parameter ε. In particular, from this result we conclude *that the error, measured as deviation from the solution of the corresponding index-3 DAE, can be estimated like*

$$\|\boldsymbol{q}_n^\varepsilon - \boldsymbol{q}(t_n)\| + \|\boldsymbol{v}_n^\varepsilon - \dot{\boldsymbol{q}}(t_n)\| \le C(h\varrho^n + h^{p_{\mathrm{DAE3}}} + \varepsilon^2)\,,$$

where p_{DAE3} *denotes the order (for position and velocity coordinates) of the method when applied to the index-3 problem* (2.12)*. There are various convergence results from the literature (Hairer et al., 1989a, Lubich, 1993, among others) proving that order reduction occurs even for families of stiffly accurate collocation methods which are commonly reckoned as the most robust and reliable methods.*

The proof of the main result (3.26) *is based on methods from (Hairer et al., 1988): It is shown that the numerical solutions may be expanded into a series in powers of* ε^2 *as the smooth solution in* (3.19)*. Since the methods are linear, this result shows that the errors for the coefficients* $(\boldsymbol{q}^i, \dot{\boldsymbol{q}}^i)^\top$ *may be studied independently. So, convergence for* (3.15) *is directly related to convergence for DAE systems of high index* 3, 5, 7, …*, see Example 3.17.*

Scholz (1989) and Simeon (1998) use the test equation of Prothero–Robinson where the leading stiff terms are only linear to analyze Runge–Kutta and Rosenbrock methods for stiff mechanical systems. The error analysis is substantially simplified and reveals that local errors *dominate the numerical behavior of the methods since other error sources get damped out by the L-stable methods. From that observation it becomes evident that the stage order plays such an important role in the above estimates.*

Schneider (1995) extended, almost without any changes, the results to a class of multistep Runge–Kutta methods and pointed out that the I-stability can even be relaxed to so-called I_d*-stability, i. e., stability on the imaginary axis for all* z *with* $|z| > d$, $d > 0$*. Finally, note that as for strongly damped systems we get a condition of the form* $0 < \varepsilon < \tilde{C}h$ *for some constant* $\tilde{C}$ *for the above estimates to hold.*

Remark 3.21 (Mixed formulation)
An obvious question is whether one should consider using both kinds of singular regularizing force terms, i. e., the system

$$\mathbf{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} = \boldsymbol{f}(\boldsymbol{q}, \dot{\boldsymbol{q}}) - \frac{1}{\delta}\mathbf{G}^\top(\boldsymbol{q})\mathbf{G}(\boldsymbol{q})\dot{\boldsymbol{q}} - \frac{1}{\varepsilon^2}\mathbf{G}^\top(\boldsymbol{q})\boldsymbol{g}(\boldsymbol{q})\,, \tag{3.27}$$

comprising stiff (spring-element) terms as well as strongly attractive damping. Kurdila et al. (1993) prove convergence and stability for this formulation, in case that without penalizing potential or dissipation the system is conservative, by the construction of a suitable Lyapunov function. Usually, (3.27) *is only used as a stabilization of the stiff or strongly attractive system to either stabilize the time integration itself since the highly oscillatory terms in* (3.15) *are difficult to tackle or to avoid drift-off in case of an integration of* (3.6) *(Hans, 2004).*

Going even one step further one might as well incorporate the constraints on acceleration level (2.14) *and include a corresponding inertia penalty by a adding a so-called 'fictitious kinetic energy term' to the Lagrangian in the system (Bayo et al., 1988). The resulting equations of motion then read*

$$\left(\left[\mathbf{M} + \tfrac{1}{\zeta}\mathbf{G}^\top\mathbf{G}\right](\boldsymbol{q})\right)\ddot{\boldsymbol{q}} = \boldsymbol{f}(\boldsymbol{q}, \dot{\boldsymbol{q}}) - \mathbf{G}^\top(\boldsymbol{q})\left(\frac{1}{\varepsilon^2}\boldsymbol{g}(\boldsymbol{q}) + \frac{1}{\delta}\mathbf{G}(\boldsymbol{q})\dot{\boldsymbol{q}} + \frac{1}{\zeta}\mathsf{R}(\boldsymbol{q})(\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}})\right)\,, \tag{3.28}$$

with yet another penalty constant $\zeta > 0$*, but the numerical challenges as well as problems like drift-off (for relatively large* ε*), numerical instability due to the ill-posed nature of the problems and the lack of a suitable way of choosing the parameters appropriately remain. Note the connection to the regularization methods from DAE time integration as Gear–Gupta–Leimkuhler* *and Baumgarte formulation. A thorough analysis for this two or even three penalty parameters appears to be rather complicated. So, mostly one parameter is supposed to cause the singular perturbation and the others are regarded as stabilizations. For the case of weak damping and Hamiltonian systems, see for instance the work of Modin and Söderlind (2011). In case of stabilization of stiff systems one typically chooses* $1/\delta := 2\xi\varepsilon^{-1}$ *with a parameter* $0 \leq \xi < 1$ *to control the amount of damping, cf. Remark 4.18 below. Kurdila et al. (1993) propose to include even further parameters to scale the constraints in order to have control on the eigenvalues. An also frequently used approach is the* augmented Lagrangian *method where constraints and stiff potentials are both used for computational advantages.*

At last, although beyond the scope of the present work, we also mention that in mechanical engineering and biomechanics it is a common approach to approximate impact models *by locally active force laws of spring-damper type. Here, many models are known in the literature and provide suitable parameter choices based on material laws (Hertz damping, Hunt–Crossley impact, Kelvin–Voigt material, etc.).*

Chapter 4

The Newmark time integration family

In this chapter we consider the general second order differential-algebraic initial value problem

$$\begin{aligned} \mathbf{M}(\boldsymbol{q}(t))\ddot{\boldsymbol{q}}(t) &= \boldsymbol{F}(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t),\boldsymbol{\lambda}(t))\,, \\ \mathbf{0} &= \boldsymbol{\Phi}(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t),\boldsymbol{\lambda}(t))\,, \quad \boldsymbol{q}(t_0)=\boldsymbol{q}_0,\ \dot{\boldsymbol{q}}(t_0)=\dot{\boldsymbol{q}}_0,\ (t\in[t_0,t_{\text{end}}]) \end{aligned} \tag{4.1}$$

with regular mass matrix $\mathbf{M}\colon\ \mathbb{R}^{n_{\boldsymbol{q}}}\to\mathbb{R}^{n_{\boldsymbol{q}}\times n_{\boldsymbol{q}}}$, generalized force vector $\boldsymbol{F}\colon\ \mathbb{R}^{2n_{\boldsymbol{q}}+n_{\boldsymbol{\lambda}}}\to\mathbb{R}^{n_{\boldsymbol{q}}}$ and constraint function $\boldsymbol{\Phi}\colon\ \mathbb{R}^{2n_{\boldsymbol{q}}+n_{\boldsymbol{\lambda}}}\to\mathbb{R}^{n_{\boldsymbol{\lambda}}}$ which are, at first, all assumed to be sufficiently smooth. For the ODE-case, $n_{\boldsymbol{\lambda}}$ might be zero. In this general form we can cope with the index-1, 2 and 3 formulations of Chapters 2 and 3 (and could even with nonholonomic constraints or friction forces) in one unified framework.

4.1 The algorithm

In the numerical solution procedure we acquire approximations $\boldsymbol{q}_n$, $\boldsymbol{v}_n$, $\boldsymbol{\lambda}_n$ for the position coordinates $\boldsymbol{q}(t_n)$, the generalized velocities $\dot{\boldsymbol{q}}(t_n)$ and the Lagrange multipliers $\boldsymbol{\lambda}(t_n)$ on an equidistant time grid $\{t_n\}_{n=0}^N$, $t_n = t_0 + nh$, $n = 0,1,\ldots,N$, $h := (t_{\text{end}} - t_0)/N$, as well as accelerations $\dot{\boldsymbol{v}}_n \approx \ddot{\boldsymbol{q}}(t_n)$ via the recursion formulae

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + h\boldsymbol{v}_n + h^2(\tfrac{1}{2}-\beta)\boldsymbol{a}_n + h^2\beta\boldsymbol{a}_{n+1}\,, \tag{4.2a}$$

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n + h(1-\gamma)\boldsymbol{a}_n + h\gamma\boldsymbol{a}_{n+1}\,, \tag{4.2b}$$

$$(1-\alpha_m)\boldsymbol{a}_{n+1} + \alpha_m\boldsymbol{a}_n = (1-\alpha_f)\dot{\boldsymbol{v}}_{n+1} + \alpha_f\dot{\boldsymbol{v}}_n\,, \tag{4.2c}$$

$$\left.\begin{aligned} \mathbf{M}(\boldsymbol{q}_{n+1})\dot{\boldsymbol{v}}_{n+1} &= \boldsymbol{F}(\boldsymbol{q}_{n+1},\boldsymbol{v}_{n+1},\boldsymbol{\lambda}_{n+1})\,, \\ \mathbf{0} &= \boldsymbol{\Phi}(\boldsymbol{q}_{n+1},\boldsymbol{v}_{n+1},\boldsymbol{\lambda}_{n+1})\,. \end{aligned}\right\} \tag{4.2d}$$

The first two equations are in the literature referred to as Newmark's method (Newmark, 1959, originally introduced in 1952) and they form the starting point of a huge variety of time integration methods for problems in structural dynamics as they have been proposed in the 1960's to 1990's, see (Hilber et al., 1977, Wood et al., 1981, Hoff and Pahl, 1988b) and the overview given by Hughes (1987). We adapt the terminology 'Newmark-type' here and throughout the present work. Equations (4.2a) and (4.2b) may simply be gained by using a combination of the explicit Euler scheme and a θ-method in the *acceleration-like variables* $\boldsymbol{a}_n \in \mathbb{R}^{n_{\boldsymbol{q}}}$. We refer to them as 'acceleration-like' because they are, as will become clear below, only a low order approximation to the actual acceleration $\ddot{\boldsymbol{q}}(t_n)$ coupled through equating two linear combinations in (4.2c). Note that the introduction of $\boldsymbol{a}_n$, $n \geq 0$, makes (4.2) a *multistep method*, although it allows for an easy onestep representation and—more importantly—a onestep-like implementation, see Remark 4.2 below.

As given in (4.2), the algorithm has four free parameters. Apart from the classical Newmark scheme, see Remark 4.1 below, all methods of practical relevance are of second order, which we will see in Section 5.1 to be generally the case, as long as the *second order condition*

$$\gamma = \frac{1}{2} - \Delta_\alpha \quad \text{for } \Delta_\alpha := \alpha_m - \alpha_f \tag{4.3}$$

is fulfilled. The probably most common parameter choice, the *Chung–Hulbert algorithm (CH(ϱ_∞) method, generalized-α method)*

$$\alpha_m = \frac{2\varrho_\infty - 1}{\varrho_\infty + 1}, \quad \alpha_f = \frac{\varrho_\infty}{\varrho_\infty + 1}, \quad \beta = \frac{1}{4}\left(\frac{1}{2} + \gamma\right)^2 = \frac{1}{(\varrho_\infty + 1)^2} \tag{4.4}$$

due to Chung and Hulbert (1993) is a one-parameter family with $\varrho_\infty \in [0, 1]$, called the *numerical damping parameter*. Note that in literature sometimes the roles of α_m and $1 - \alpha_m$ as well as α_f and $1 - \alpha_f$ are interchanged (Jansen et al., 2000, Rang, 2013) leading to the parameters $\alpha_m = \frac{2-\varrho_\infty}{1+\varrho_\infty}$, $\alpha_f = \frac{1}{1+\varrho_\infty}$, $\beta = \frac{1}{4}(1 + \alpha_m - \alpha_f)^2$, $\gamma = \frac{1}{2} - \alpha_f + \alpha_m$. The methods are nevertheless equivalent.

Remark 4.1 (The classical Newmark scheme: Derivation and prominent special cases)
In the original work, Newmark (1959) introduces solely the position and velocity updates (4.2a) *and* (4.2b) *which one obtains as a special case of the* (4.2) *for $\alpha_m = \alpha_f$. For these parameters the acceleration-like variables $\boldsymbol{a}_n$ become unnecessary since they coincide with $\dot{\boldsymbol{v}}_n$ not taking into account the initialization procedure. The original development of the method is based on a representation of $\boldsymbol{q}(t)$ in the form*

$$\underbrace{\boldsymbol{q}(t_{n+1})}_{\approx \boldsymbol{q}_{n+1}} = \underbrace{\boldsymbol{q}(t_n)}_{\approx \boldsymbol{q}_n} + h\,\underbrace{\dot{\boldsymbol{q}}(t_n)}_{\approx \boldsymbol{v}_n} + \int_{t_n}^{t_{n+1}} (t_{n+1} - \tau)\,\underbrace{\ddot{\boldsymbol{q}}(\tau)}_{\approx \bar{\boldsymbol{a}}(\tau)}\,\mathrm{d}\tau\,, \quad \underbrace{\dot{\boldsymbol{q}}(t_{n+1})}_{\approx \boldsymbol{v}_{n+1}} = \underbrace{\dot{\boldsymbol{q}}(t_n)}_{\approx \boldsymbol{v}_n} + \int_{t_n}^{t_{n+1}} \underbrace{\ddot{\boldsymbol{q}}(\tau)}_{\approx \hat{\boldsymbol{a}}(\tau)}\,\mathrm{d}\tau\,, \tag{4.5}$$

where the functions $\bar{\boldsymbol{a}}$, $\hat{\boldsymbol{a}}\colon [t_n, t_{n+1}] \to \mathbb{R}^{n_q}$ serve as non-discrete approximations of $\ddot{\boldsymbol{q}}(\tau)$. Using the approximation $\bar{\boldsymbol{a}}(\tau) = \hat{\boldsymbol{a}}(\tau) \equiv \boldsymbol{a}_n$ and carrying out the integration in (4.5) *leads to the explicit Euler method with $\gamma = \beta = 0$. Equivalently, $\bar{\boldsymbol{a}} = \hat{\boldsymbol{a}} = \boldsymbol{a}_{n+1}$, and so $\gamma = 1$, $\beta = \frac{1}{2}$ is the implicit Euler method.*

All three approximations with $\bar{\boldsymbol{a}} = \hat{\boldsymbol{a}}$ and

a) $\bar{\boldsymbol{a}}(\tau) = \begin{cases} \boldsymbol{a}_n & \tau \in [t_n, t_n + h/2] \\ \boldsymbol{a}_{n+1} & \textit{else} \end{cases}$

b) $\bar{\boldsymbol{a}}(\tau) = \boldsymbol{a}_n + \dfrac{(\tau - t_n)(\boldsymbol{a}_{n+1} - \boldsymbol{a}_n)}{h}$

c) $\bar{\boldsymbol{a}}(\tau) \equiv \dfrac{\boldsymbol{a}_n + \boldsymbol{a}_{n+1}}{2}$

*lead to $\gamma = \frac{1}{2}$ and a) $\beta = \frac{1}{8}$ (step function approximation), b) $\beta = \frac{1}{6}$ (*linear acceleration method*), and c) $\beta = \frac{1}{4}$ (*constant acceleration method, trapezoidal rule, Verlet scheme *for unconstrained systems and $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{q})$,* RATTLE *in the constrained case). Keeping $\bar{\boldsymbol{a}}$ as above and $\hat{\boldsymbol{a}}(\tau) \equiv \boldsymbol{a}_n$ is equivalent to $\beta = 0$ and commonly known as* Störmer's method *which for velocity-independent $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{q})$ then coincides with the* explicit *central difference scheme. At last, $\gamma = \frac{1}{2}$, $\beta = \frac{1}{12}$ is known as* Fox–Goodwin *scheme or royal-road method and designed to minimize period errors (Fox and Goodwin, 1949). 'The' Newmark method (or average constant acceleration method) is a family with parameter choices $\gamma \in [\frac{1}{2}, 1]$, $\beta = \frac{1}{4}(\frac{1}{2} + \gamma)^2$, see Example 4.17 below. The* *case $\gamma = \frac{1}{2}$ plays an important role because only in that setting the methods are second order accurate. The reason to even consider implicit methods that are only of first order lies in the improved stability behavior or* numerical damping, *which will become evident in Section 4.2.*

The algorithm in the above general form, i. e., with the four parameters β, γ, α_m, and α_f has first been introduced by Chung and Hulbert (1993) as a special case of the six-parameter algorithm of Hoff and Pahl (1988a). The derivation of that algorithm was based on a subspace collocation approach adapting methods from finite-element analysis (moment-matching for an arbitrary weighting function) within the framework introduced by Zienkiewicz (1977).

Remark 4.2 (Characterization as a multistep method)
Following the idea of Erlicher et al. (2002) we state the above algorithm in the form of a partitioned linear multistep method. The position and velocity updates (4.2a) *and* (4.2b) *on two consecutive time steps allow for an elimination of the acceleration-like variables $\boldsymbol{a}_n$ such that we obtain a three-level recursion that only includes position and velocity variables. In the same way (see Arnold and Brüls, 2007) we can use* (4.2b) *and* (4.2c) *to eliminate $\boldsymbol{a}_{n-1}$, $\boldsymbol{a}_n$ and $\boldsymbol{a}_{n+1}$ such that we arrive at the relations*

$$\begin{aligned} \sum_{i=0}^{2} \alpha_i^{\boldsymbol{q}} \boldsymbol{q}_{n+i-1} &= h \sum_{i=0}^{2} \beta_i^{\boldsymbol{v}} \boldsymbol{v}_{n+i-1} \,, \\ \sum_{i=0}^{2} \alpha_i^{\boldsymbol{v}} \boldsymbol{v}_{n+i-1} &= h \sum_{i=0}^{2} \beta_i^{\boldsymbol{v}} \dot{\boldsymbol{v}}_{n+i-1} \end{aligned} \tag{4.6}$$

with

$$\begin{aligned} \alpha_0^{\boldsymbol{q}} &= 2\gamma - 1\,, & \alpha_1^{\boldsymbol{q}} &= \gamma - 1\,, & \alpha_2^{\boldsymbol{q}} &= \gamma\,, \\ \beta_0^{\boldsymbol{q}} &= \frac{1 + 2\beta - 2\gamma}{2}\,, & \beta_1^{\boldsymbol{q}} &= \frac{1 - 4\beta + 2\gamma}{2}\,, & \beta_2^{\boldsymbol{q}} &= \beta\,, \\ \alpha_0^{\boldsymbol{v}} &= -\alpha_m\,, & \alpha_1^{\boldsymbol{v}} &= 2\alpha_m - 1\,, & \alpha_2^{\boldsymbol{v}} &= 1 - \alpha_m\,, \\ \beta_0^{\boldsymbol{v}} &= \alpha_f(1-\gamma)\,, & \beta_1^{\boldsymbol{v}} &= 1 - \gamma + \alpha_f(2\gamma - 1)\,, & \beta_2^{\boldsymbol{v}} &= \gamma(1-\alpha_f)\,. \end{aligned}$$

Additionally, the equilibrium conditions (4.2d) *remain valid and define the acceleration variables $\dot{\boldsymbol{v}}_{n+1}$ and Lagrange multiplier vectors $\boldsymbol{\lambda}_{n+1}$ while also ensuring constraint fulfillment. So, a comparison with* (2.29) *shows that algorithm* (4.2) *falls into the class of partitioned linear two-step methods. Note that when given as in* (4.6) *within the initialization of the algorithm there is an additional degree of freedom because instead of just setting $\boldsymbol{a}_0$ one has to compute values for $\boldsymbol{q}_1$ and $\boldsymbol{v}_1$ to start the algorithm. Note also that in the literature the classification of the algorithm is handled differently as it combines properties of onestep as well as multistep methods. Hughes (1987) uses the term 'onestep-multivalue method' to characterize Newmark integrators.*

Remark 4.3 (Extension to nonlinear systems)
Originally the above method has been introduced for linear systems

$$\mathbf{M}\ddot{\boldsymbol{q}}(t) + \mathbf{D}\dot{\boldsymbol{q}}(t) + \mathbf{K}\boldsymbol{q}(t) = \boldsymbol{f}(t)\,, \tag{4.7}$$

where only the load vector $\boldsymbol{f}$ *explicitly depends on time. These systems usually appear in large scale semi-discretized finite-element simulations in structural dynamics, cf. Example 3.19. In their original work, Chung and Hulbert (1993) used a different formulation imposing the equi-*

librium condition, also called collocation or balance equation, at a shifted time instance $t_{n+1-\alpha_f}$.

$$\begin{array}{rcll} & \mathbf{M}\boldsymbol{a}_{n+1-\alpha_m} + \mathbf{D}\boldsymbol{v}_{n+1-\alpha_f} + \mathbf{K}\boldsymbol{q}_{n+1-\alpha_f} = \boldsymbol{f}(t_{n+1-\alpha_f})\,, & & \\ \textit{where}\quad t_{n+1-\alpha_f} & := & \alpha_f t_n + (1-\alpha_f)t_{n+1}\,, & \\ \boldsymbol{q}_{n+1-\alpha_f} & := & \alpha_f \boldsymbol{q}_n + (1-\alpha_f)\boldsymbol{q}_{n+1} & \approx\ \boldsymbol{q}(t_{n+1-\alpha_f})\,, \\ \boldsymbol{v}_{n+1-\alpha_f} & := & \alpha_f \boldsymbol{v}_n + (1-\alpha_f)\boldsymbol{v}_{n+1} & \approx\ \dot{\boldsymbol{q}}(t_{n+1-\alpha_f})\,, \\ \boldsymbol{a}_{n+1-\alpha_m} & := & \alpha_m \boldsymbol{a}_n + (1-\alpha_m)\boldsymbol{a}_{n+1} & \approx\ \ddot{\boldsymbol{q}}(t_{n+1-\alpha_f})\,. \end{array} \tag{4.8}$$

Note that the indices are just notations; only variables with integer-subscripts are relevant for later use. The linear structure of (4.8) *leaves some freedom when extending the algorithm to general nonlinear systems. A direct transition to nonlinear systems—and the way it has originally been proposed by Hulbert and Chung (1996), see also (Hilber and Hughes, 1978, Erlicher et al., 2002, Rang, 2013)—would lead to an algorithm where the equilibrium condition* (4.2d) *is to be replaced by*

$$\begin{aligned} \mathbf{M}(\boldsymbol{q}_{n+1-\alpha_f})\boldsymbol{a}_{n+1-\alpha_m} &= \boldsymbol{F}(\boldsymbol{q}_{n+1-\alpha_f}, \boldsymbol{v}_{n+1-\alpha_f}, \boldsymbol{\lambda}_{n+1-\alpha_f})\,, \\ \mathbf{0} &= \boldsymbol{\Phi}(\boldsymbol{q}_{n+1-\alpha_f}, \boldsymbol{v}_{n+1-\alpha_f}, \boldsymbol{\lambda}_{n+1-\alpha_f})\,, \end{aligned} \tag{4.9}$$

and $\boldsymbol{\lambda}_{n+1-\alpha_f} := \alpha_f \boldsymbol{\lambda}_n + (1-\alpha_f)\boldsymbol{\lambda}_{n+1}$. *Note that the (true) acceleration variables* $\dot{\boldsymbol{v}}_n$ *are no longer present in this form and are typically not computed. We call* (4.8) *and* (4.9) *the* one-leg version *of Newmark integrators (Hairer and Wanner, 2002) or* midpoint collocation *with regards to the work of Hilber and Hughes (1978). Note, in particular, that for linear systems, trapezoidal rule and implicit midpoint rule lead to the same algorithm.*

Another way of extending to nonlinear systems and implementing the algorithms is given if the acceleration update (4.2c) *is plugged into the equilibrium condition, i. e.,*

$$\begin{aligned} (1-\alpha_m)\mathbf{M}\boldsymbol{a}_{n+1} + \alpha_m \mathbf{M}\boldsymbol{a}_n &= (1-\alpha_f)\boldsymbol{F}(\boldsymbol{q}_{n+1}, \boldsymbol{v}_{n+1}, \boldsymbol{\lambda}_{n+1}) + \alpha_f \boldsymbol{F}(\boldsymbol{q}_n, \boldsymbol{v}_n, \boldsymbol{\lambda}_n)\,, \\ \mathbf{0} &= (1-\alpha_f)\boldsymbol{\Phi}(\boldsymbol{q}_{n+1}, \boldsymbol{v}_{n+1}, \boldsymbol{\lambda}_{n+1}) + \alpha_f \boldsymbol{\Phi}(\boldsymbol{q}_n, \boldsymbol{v}_n, \boldsymbol{\lambda}_n)\,, \end{aligned} \tag{4.10}$$

called the modified residual equations *(Brüls and Golinval, 2006) as is preferred among others by Lunk and Simeon (2006) and Jay and Negrut (2007).*

For nonconstant mass matrix it is not obvious which arguments $\boldsymbol{q}$ *of* $\mathbf{M}$ *are to be taken in the left-hand-side of* (4.10). *Simply inserting* $\boldsymbol{q}_{(n-1)+1-\alpha_f}$ *and* $\boldsymbol{q}_{n+1-\alpha_f}$ *would make the computations dependent on* $\boldsymbol{q}_{n-1}$ *which would destroy the beneficial onestep implementation structure, see Remark 4.19 below. Jay (2011) proposed to calculate the arguments* $\boldsymbol{q}$ *as explicit Euler predictors using only* $\boldsymbol{q}_n$ *and* $\boldsymbol{v}_n$. *For the HHT method to be defined in Example 4.17 below we have* $\alpha_m = 0$ *and this is no issue. On the other hand, for* $\alpha_f = 0$, *which is commonly known as WBZ or Bossak–Newmark algorithm, see below, the equilibrium condition for all formulations is to be taken at* t_{n+1} *and* (4.2), (4.9), *and* (4.10) *coincide for constant* $\mathbf{M}$. *Note that all algorithms in this remark are just* $\mathcal{O}(h^2)$*-perturbations of each other. For nonstiff nonlinear ODE systems the convergence results, for instance of Erlicher et al. (2002), are valid for all these formulations. We also refer to (Brüls, 2005, Section 3.3.2) and (Géradin and Rixen, 2015, Remark 7.3) for a discussion of the different formulations.*

Taking the collocation point for the equilibrium at $t = t_{n+1}$ *is advantageous for many reasons: (a) Nonconstant mass matrices are easily and straightforwardly incorporated, (b) constraint enforcement for all approximations on the time grid is a reasonable requirement, (c) for general mechatronic systems with control feedback, exact knowledge of the accelerations on the time grid is very important (Brüls and Arnold, 2008), the same is true for the forces of models in sustainability analysis, (d) for some large scale PDE or control problems it is not easy to obtain force terms at intermediate values, and (e) as BDF methods using a residual formalism (Arnold* *et al., 2011) are the most common implicit integrators in multibody dynamics, the adaptation of existing code for the use of Newmark-type integrators seems easier using the approach favored in this thesis.*

Remark 4.4 (Industrial implementations)
As they have originally been introduced for problems of structural dynamics, Newmark-type integrators are usually available for commercial simulation environments for partial differential equations. The simulation platform ComSol-Multiphysics (COMSOL (2008), COMSOL (2012)) offers a Chung–Hulbert(ϱ_∞) algorithm with local error estimation for the simulation of dynamical models. Newmark integrators, as variable time step HHT methods, see Example 4.17 below, are also implemented for the time integration of large scale finite element models in the finite-element tool Abaqus (SIMULIA (2011)). The MSC software collection uses an averaged version of the classical Newmark method for large finite-element transient analysis (MSC-Nastran) and Newmark and HHT in modified residual form for simulations in MSC-Adams, see (MSC-Software (2012)) and the description by Negrut et al. (2005). LS-DYNA (Livermore Software Tech.-Corp., 2006/2007) 'implicit-dynamics' computations also use Newmark as the default time integration scheme.

In the Matlab-based multibody simulation tool NEWEUL-M^2 (Kurz et al., 2010) there are variable step size implementations of the classical Newmark scheme as well as HHT and Chung–Hulbert(ϱ_∞) method. The first multibody simulation environment using the version of Newmark methods for DAEs as they are presented here is the CAE (computer aided engineering) software MECANO (Samtech (2015), Brüls (2005)). The simulation platform RecurDyn (RecurDyn (2015)) also offers the possibility to use a Chung–Hulbert(ϱ_∞) method for computational models. Its 'Hybrid Integrator' is a Newmark integration scheme with the parameters β and γ chosen as for the Chung–Hulbert(ϱ_∞) method (Sanborn et al., 2014). The free simulation environment FreeDyn (Nachbagauer et al., 2015) uses a variant of HHT as it is proposed by Negrut et al. (2005). The multibody simulation tool Universal Mechanism (Universal Mechanism (2015)) uses a variation of Park's method (Fung and Tong, 2001) which is closely related to the 'classical' integration scheme of Newmark.

4.2 Linear stability analysis and optimal parameter choice

4.2.1 Linear test equation and stability

The study of stability for algorithms designed for ordinary differential equations is inevitably linked to the famous test equation

$$\dot{x}(t) = \lambda x(t)\,,\ x(0) = x_0\,, \quad \Re\lambda \leq 0,\ , \tag{4.11}$$

of Dahlquist (1963). As the norm of the analytic solution $x(t) = x_0\,\mathrm{e}^{-\lambda t}$ of (4.11) is non-increasing, linear stability of a numerical procedure is defined as the property of providing bounded solutions for the application to (4.11). Typically, this leads to restrictions on the product $\tilde{z} := h\lambda$ of step size and stiffness parameter. For Runge–Kutta methods, one step is given by $x_1 = R(\tilde{z}) \cdot x_0$ with the stability function $R(\tilde{z})$ such that the stability region S indeed describes the region of all $h\lambda$ giving stable solutions in the above sense.

For second order differential equations the test equation is also due to Dahlquist (1978) and given by the harmonic oscillator (3.2)

$$\ddot{q}(t) + \omega^2 q(t) = 0\,,\ q(0) = 0\,,\ \dot{q}(0) = \dot{q}_0\,, \tag{4.12}$$

where the singular perturbation parameter is replaced by its inverse $\omega = \varepsilon^{-1}$. The analytic solutions are of the form

$$q(t) = C_1 \sin(\omega t) + C_2 \cos(\omega t) \tag{4.13}$$

with C_1, $C_2 \in \mathbb{R}$ depending on the initial data. Again stability is by definition given if the approximations remain bounded; we will introduce more detailed stability concepts in Definition 4.7. The consideration of (4.11) and (4.12) is motivated by the fact that in a neighborhood of equilibrium points both test equations resemble the leading term of any ODE after local linearization and decomposition. We will see in Chapter 5 that this classical consideration already covers the governing linear part of the analysis for the singularly perturbed mechanical systems.

Gladwell and Thomas (1980) propose to consider the test equation

$$\ddot{q}(t) + 2\xi\dot{q}(t) + (\xi^2 + \omega^2)q(t) = 0 \tag{4.14}$$

in order to relate the second order integration methods directly to those for first order equations. Its analytic solution comprises terms of the form $\mathrm{e}^{-\xi t} \cdot \sin(\omega t)$ and $\mathrm{e}^{-\xi t} \cdot \cos(\omega t)$ and therefore resembles the numerical behavior for damped oscillations as in the scalar case (4.12) with $\lambda = \xi \pm \mathrm{i}\,\omega$. This approach allows the definition of stability regions as for first order equations: The stability region for (4.2) comprises all points $(h\xi_0, h\omega_0) \in \mathbb{R}^2$ such that the algorithm provides stable solutions for (4.14) with $\xi = \xi_0$, $\omega = \omega_0$, and step size $h \geq 0$.

A straightforward application of the Newmark integrator (4.2) to (4.12) leads to the linear recursion formula (Chung and Hulbert, 1993)

$$\begin{pmatrix} 1 & 0 & -\beta \\ 0 & 1 & -\gamma \\ (1-\alpha_f)z^2 & 0 & 1-\alpha_m \end{pmatrix} \begin{pmatrix} q_{n+1} \\ hv_{n+1} \\ h^2 a_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0.5-\beta \\ 0 & 1 & 1-\gamma \\ -\alpha_f z^2 & 0 & -\alpha_m \end{pmatrix} \begin{pmatrix} q_n \\ hv_n \\ h^2 a_n \end{pmatrix}, \tag{4.15}$$

where the variable $z := h\omega$ has been introduced. For later reference the linear operator of the recursion is denoted by $\mathbf{T}(z)$.

$$\begin{aligned} \begin{pmatrix} q_n & hv_n & h^2 a_n \end{pmatrix}^\top &\rightsquigarrow \begin{pmatrix} q_{n+1} & hv_{n+1} & h^2 a_{n+1} \end{pmatrix}^\top = \mathbf{T}(z) \begin{pmatrix} q_n & hv_n & h^2 a_n \end{pmatrix}^\top, \\ \mathbf{T}(z) &= \begin{pmatrix} 1 & 0 & -\beta \\ 0 & 1 & -\gamma \\ (1-\alpha_f)z^2 & 0 & 1-\alpha_m \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 0.5-\beta \\ 0 & 1 & 1-\gamma \\ -\alpha_f z^2 & 0 & -\alpha_m \end{pmatrix} \\ &= \begin{pmatrix} \dfrac{\alpha_f \beta z^2 + \alpha_m - 1}{(\alpha_f - 1)\beta z^2 + \alpha_m - 1} & \dfrac{\alpha_m - 1}{(\alpha_f - 1)\beta z^2 + \alpha_m - 1} & \dfrac{\alpha_m + 2\beta - 1}{2\left((\alpha_f - 1)\beta z^2 + \alpha_m - 1\right)} \\ \dfrac{z^2\gamma}{(\alpha_f - 1)\beta z^2 + \alpha_m - 1} & 1 - \dfrac{z^2(\alpha_f - 1)\gamma}{(\alpha_f - 1)\beta z^2 + \alpha_m - 1} & \dfrac{0.5\left(2 - z^2(\alpha_f - 1)\right)\gamma}{\left((\alpha_f - 1)\beta z^2 + \alpha_m - 1\right)} + 1 \\ \dfrac{z^2}{(\alpha_f - 1)\beta z^2 + \alpha_m - 1} & -\dfrac{z^2(\alpha_f - 1)}{(\alpha_f - 1)\beta z^2 + \alpha_m - 1} & \dfrac{(\alpha_f - 1)(2\beta - 1)z^2 + 2\alpha_m}{2\left((\alpha_f - 1)\beta z^2 + \alpha_m - 1\right)} \end{pmatrix}, \end{aligned} \tag{4.16}$$

where we assume β, $\gamma \neq 0$, α_m, $\alpha_f \neq 1$ as will become evident in the analysis below. To study the error growth, eigenvalues of $\mathbf{T}(z)$ need to be estimated because they govern the behavior of (4.15) for repeated application. As the entries of the three-by-three update formula depend on five parameters, the calculations are very involved. Following Brüls (2005), a condensed way of providing the characteristic polynomial is given by $\chi_{\mathbf{T}}(\mu) = \left[P_{\alpha_f}(P_\gamma + P_1 P_\beta)z^2 + P_1^2 P_{\alpha_m}\right](\mu)$, where

$$\begin{aligned} P_{\alpha_m}(\mu) &= (1-\alpha_m)\mu + \alpha_m\,, \qquad & P_{\alpha_f}(\mu) &= (1-\alpha_f)\mu + \alpha_f\,, \qquad & P_\gamma(\mu) &= \gamma\mu + 1 - \gamma\,, \\ P_\beta(\mu) &= \beta\mu + \tfrac{1}{2} - \beta\,, & P_1(\mu) &= \mu - 1\,. \end{aligned}$$

Before we start the stability analysis we will have a closer look at the generalized-α algorithm.

Example 4.5 (Stability of the CH(ϱ_∞) algorithm)
To illustrate the stability—more precisely: the damping properties—of the numerical integration schemes, it has become common practice to plot the maximum absolute value of the eigenvalues of the amplification matrix $\mathbf{T} = \mathbf{T}(z)$ as a function of $z = h\omega$. In Figure 4.1 we illustrate this numerical damping behavior for the CH(ϱ_∞) method (4.3), (4.4) using different values of the parameter ϱ_∞. Now the reference to ϱ_∞ as 'numerical damping' becomes clear as this parameter controls the amount of damping in the high frequency regime. Judging from this illustration, it would appear reasonable to prefer $\varrho_\infty \approx 1$ because $q(t)$ in (4.13) is periodic, i. e., its local minima/maxima do not decrease in absolute value over time. Nevertheless, in practical implementations and for highly nonlinear or large scale problems the lack of numerical damping of CH(ϱ_∞) with $\varrho_\infty \approx 1$ leads to undesired amplification of oscillations and therefore nonphysical behavior of the numerical approximations such that numerical damping becomes a desirable feature. We will come back to this topic in Remark 4.20. Note that from the definition of z it follows that larger step sizes not only imply a more effective solution, as less integration steps are needed, but also more damping. ♢

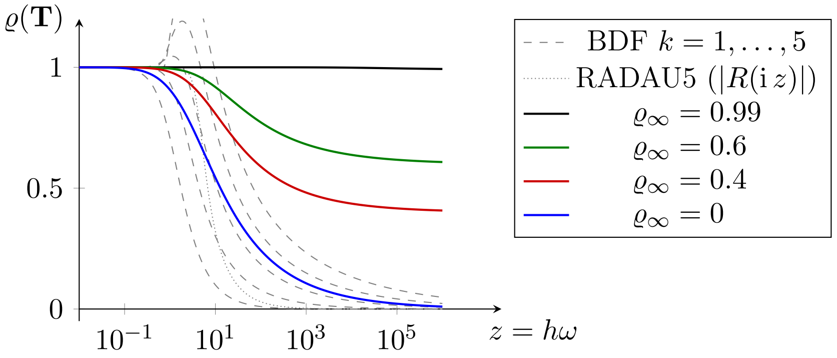


Figure 4.1: Spectral radii of amplification matrices for varying values of ϱ_∞

Remark 4.6 (Alternative scaling)
Studying the relative growth in $(q_n, hv_n, h^2 a_n)^\top$ is the 'classical way' of analyzing the stability of integration methods from structural dynamics and appears natural since all involved quantities have the same physical units. Of course, any rescaling of these variables has no influence on the asymptotic behavior. For the analysis of constrained mechanical systems, it is convenient to rescale by ω^2 and h^{-2}, respectively, so Arnold et al. (2016) propose the mapping

$$\begin{pmatrix} z^{-2} & 0 & -\beta \\ 0 & 1 & -\gamma \\ 1-\alpha_f & 0 & 1-\alpha_m \end{pmatrix} \begin{pmatrix} \omega^2 q_{n+1} \\ \frac{1}{h} v_{n+1} \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} z^{-2} & 1 & \frac{1}{2}-\beta \\ 0 & 1 & 1-\gamma \\ -\alpha_f & 0 & -\alpha_m \end{pmatrix} \begin{pmatrix} \omega^2 q_n \\ \frac{1}{h} v_n \\ a_n \end{pmatrix} . \tag{4.17}$$

Note that this only results in a similarity transformation of the iteration matrix $\mathbf{T}(z)$, the spectrum remains unchanged. Nevertheless, for some choices of scaling the relations, entries of $\mathbf{T}(z)$ might become unbounded for $z \to \infty$ which is an important limit case in the analysis.

The acceleration-like variables $\boldsymbol{a}_n$ are just low-order approximations of $\ddot{\boldsymbol{q}}$ which may motivate to disregard them from this map. Following the multistep representation of Erlicher et al. (2002), Kettmann (2009) derives the recursion in form $(z^2 q_n, z^2 q_{n-1}, hv_{n-1})$, such that only second order values enter the recursion and the multistep character of the schemes becomes explicitly present, see the argumentation for the second order Prothero–Robinson problem in Example 5.28 below.

Definition 4.7 (Stability notions)**.**

(a) (Dahlquist (1956), see also Hairer et al. (1993)) A time integration method for second order systems is called *zero stable* (or stable for $h \to 0$) if numerical solutions of (4.12) for $\omega = 0$, in exact arithmetics, remain bounded for arbitrary initializations of the method and $h \to 0$.

(b) It is called *stable at infinity* if numerical solutions of (4.12) for $\varepsilon \to 0$ (or $\omega \to \infty$), in exact arithmetics, remain bounded for arbitrary initializations and *strictly stable at infinity* if they even tend towards zero for $n \to \infty$.

(c) (Dahlquist, 1978) The method is called *unconditionally stable* if for fixed ω the solutions remain bounded for all values of the time step size $h > 0$.

Remark 4.8

(a) In honor of Dahlquist, zero stability is sometimes also referred to as D-stability or—in the literature on Newmark-type methods—just 'stability'. The necessity for zero stability of the integrators reflects them being multistep methods, cf. Remark 4.2, as for onestep methods zero stability is no part of the numerical analysis.

(b) In classical textbooks (e. g. Hairer et al., 1993) the above stability notions are introduced by means of so-called ρ- and σ-polynomials and a corresponding (strong) root condition. We use the above definitions as they are more intuitive.

(c) The unconditional stability for the methods under consideration is directly related to I-stability of numerical methods for first order systems. The test equation (4.14)*, as well as* (4.26) *below, allow for a definition of A- and L-stability for those methods as well. In fact, one can show that* (4.2) *is A-stable if it is unconditionally stable. Whenever solutions tend towards zero for $n \to \infty$, $\omega \neq 0$, the methods are called* absolutely stable *(Petzold et al., 1997).*

(d) Concerning the order and stability there are the two fundamental results called 'Dahlquist barriers' giving bounds for the highest attainable order of stable linear multistep methods. The first theorem is due to Dahlquist (1956) and states the maximum order of zero stable k-step methods to be $k+2$ for even k and $k+1$ if k is odd. Moreover, explicit linear multistep methods have at most order k. Methods (4.2) *are two-step such that this barrier implies a theoretical maximum order of four. More regulative is the second barrier (Dahlquist, 1963) which bounds the maximum order of an unconditionally (A-)stable method by two. This theorem has also been extended to methods for second order systems by Dahlquist (1978) and Hairer (1979). The 'best' unconditionally stable multistep method, judging from the error constant, is the trapezoidal rule.*

If the method being analyzed is stable at infinity, the absolute value of the onestep amplification mapping, i. e., the spectral radius of the amplification matrix is called *numerical damping rate*, cf. Example 4.5.

As the three-by-three structure collecting (error-) growth on position, velocity and acceleration level is prototypical for the analysis of time integration methods for mechanical problems and two eigenvalues are always close to one, the following notions have been established.

Definition 4.9 (Principal and spurious roots)**.**
Branches of complex conjugated eigenvalues of the amplification matrices in the low frequency range are called *principal roots* of (4.16). If there is a single real-valued eigenvalue for $z \to 0$, this branch is denoted as the *spurious root.*

Throughout the development of time integration methods in structural dynamics there has been a recurring discussion on the influence of the spurious root (see for example Hulbert and Chung, 1994). To gain reasonable accuracy, it is recommended that the principal roots remain complex under the Nyquist frequency. For the trapezoidal rule, i. e., (4.4) for $\varrho_\infty = 1$, the principal roots are complex and of absolute value one coinciding for $z = 0$ at $+1$ and at -1 for $z \to \infty$; the spurious root is identically -1 but of no concern at all since for all classical Newmark methods ($\alpha_m = \alpha_f$), it is possible to fully describe the discrete dynamics by means of a two-by-two recursion formula. Note that the terms are not used consistently in the literature. Sometimes (e. g. Gladwell and Thomas, 1980) principal roots are defined as those root branches that dominate the stability behavior, i. e., those of maximal absolute value, possibly in the high-frequency range.

Lemma 4.10 (Zero stability, cf. (Erlicher et al., 2002) and (Arnold and Brüls, 2007))
The methods of the Newmark time integration family (4.2) are zero stable provided that

$$\alpha_m \leq \frac{1}{2} \tag{4.18}$$

or

$$\alpha_m = \alpha_f$$

is fulfilled.

Proof. A scaled form of the onestep mapping (4.16) for $\omega = 0$ has the form

$$\begin{pmatrix} q_{n+1} \\ v_{n+1} \\ a_{n+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & h & \frac{1-\alpha_m-2\beta}{2(1-\alpha_m)}h^2 \\ 0 & 1 & \frac{1-\alpha_m-\gamma}{1-\alpha_m}h \\ 0 & 0 & \frac{-\alpha_m}{1-\alpha_m} \end{pmatrix}}_{=:\mathbf{T}_0(h)} \begin{pmatrix} q_n \\ v_n \\ a_n \end{pmatrix} . \tag{4.19}$$

In the limit case $\lim_{h\to 0} \mathbf{T}_0(h)$ becomes a diagonal matrix whose asymptotic growth under self multiplication is governed by the spurious root $\mu_3 = \frac{-\alpha_m}{1-\alpha_m}$. The norm of $(\lim_{h\to 0} \mathbf{T}_0(h))^n$, $n \geq 0$, can be estimated by one provided that $\alpha_m \leq \frac{1}{2}$ as stated. Note that for $\alpha_m = \frac{1}{2}$ it holds $\mu_3 = -1$.

For $\alpha_m = \alpha_f$, the specific value of α_m is of no concern at all since then the mapping can be expressed as $\mathbb{R}^2 \to \mathbb{R}^2$ and only the two principal eigenvalues $\mu_1 = \mu_2 = 1$ are present. □

Remark 4.11 (Zero stability)

(a) *The proofs of zero stability of Erlicher et al. (2002) and Arnold and Brüls (2007) rely on representations of the algorithms where the acceleration-like variables a_n have been eliminated. In that case, one can ensure stability by checking the strong root condition, i. e., a criterion on the multiplicity of zeros of characteristic polynomials of the respective difference equations (Hairer et al., 1993). Note that in (4.19) the mapping has an eigenvalue of absolute value one and modulus two, cf. Hairer (1979) and (Hairer et al., 1993, Definition III.10.1).*

(b) *In view of Definition 4.9, the double eigenvalue at $h\omega = z = 0$ branches for $z > 0$ into the principal roots while the spurious root-branch stems from μ_3.*

(c) *The spurious root μ_3 depends only on—and vanishes with—the parameter α_m. In the literature zero spurious roots for $h \to 0$ have been favored by many authors, see for instance the discussion of Hoff and Pahl (1988a) and may be seen as one of the major* *reasons for the popularity of the Newmark scheme and HHT, compared to WBZ. Only for $\varrho_\infty = \frac{1}{2}$, the CH($\varrho_\infty$) method also has this property which in the context of linear multistep and general linear methods is denoted as optimal zero stability (Strehmel et al., 2012). A justification will be given in Section 4.2.2.*

As the stability analysis of multistep methods reduces to the question of whether eigenvalues have absolute values greater than one and an analytic derivation of these eigenvalues is rather cumbersome, we will use the following result which allows us to give an answer without explicit knowledge of the roots.

Lemma 4.12 (Routh–Hurwitz criterion (Schwarz (1956), formulation adapted from Prüß et al. (2008)))
For the given polynomial

$$p(\zeta) = \zeta^n + a_{1,1}\zeta^{n-1} + a_{0,2}\zeta^{n-2} + a_{1,2}\zeta^{n-3} + \dots$$

with positive real-valued coefficients $a_{i,j}$, $i = 0, 1$, $j = 1, 2, \dots$, define the *Hurwitz-matrix* $(a_{i,j})_{i,j=1,\dots,n}$ via the recursion

$$a_{0,1} := 1, \quad a_{i+1,j} := \begin{cases} a_{i-1,j+1} - r_i a_{i,j+1} & \text{if } a_{i,1} \neq 0 \\ 0 & \text{else} \end{cases} \quad \text{with } r_i := \frac{a_{i-1,1}}{a_{i,1}},\ i, j > 0\,.$$

Then the following two assertions are equivalent:

(a) All roots ζ_i, $i = 1, \dots, n$, of $p(\zeta) = 0$ have negative real part.

(b) $a_{1,i} > 0$ for $i = 1, \dots, n$.

For $n = 3$, we infer that $a_3\zeta^3 + a_2\zeta^2 + a_1\zeta + a_0$ has only roots in the left halfplane iff $a_i > 0$, $i = 0, 1, 2, 3$, and $a_1a_2 > a_0a_3$.

Corollary 4.13 (Eigenvalues of a cubic polynomial (Hughes, 1987))
All roots of the cubic polynomial

$$p(\mu) = \mu^3 + b_2\mu^2 + b_1\mu + b_0\,,\ b_i \in \mathbb{R},\ i = 0, 1, 2\,,$$

lie in the interior of the unit circle if

$$\begin{aligned} &1 + b_0 + b_1 + b_2 > 0\,, \quad 3 - 3b_0 - b_1 + b_2 > 0\,, \quad 3 + 3b_0 - b_1 - b_2 > 0\,, \\ &1 - b_0 + b_1 - b_2 > 0\,, \quad \text{and } 1 + b_0b_2 > b_0^2 + b_1\,. \end{aligned} \tag{4.20}$$

Proof. We introduce the 'Greek-Roman transformation' (Hairer et al., 1993, proof of Theorem III.3.5)

$$\mu := \frac{1+\zeta}{1-\zeta}\,, \quad \zeta = \frac{\mu - 1}{\mu + 1}\,, \tag{4.21}$$

which maps the interior of the unit circle to the left halfplane, see Figure 4.2. So, $|\mu| < 1$ is equivalent to $\Re\zeta < 0$ such that Lemma 4.12 can be applied. It holds

$$\begin{aligned} &(1-\zeta)^3 \cdot p(\tfrac{1+\zeta}{1-\zeta}) \\ &= \underbrace{(1 - b_0 + b_1 - b_2)}_{=a_3}\zeta^3 + \underbrace{(3 + 3b_0 - b_1 - b_2)}_{=a_2}\zeta^2 + \underbrace{(3 - 3b_0 - b_1 + b_2)}_{=a_1}\zeta + \underbrace{(1 + b_0 + b_1 + b_2)}_{=a_0}\,, \end{aligned}$$

such that (4.20) follows from the Routh–Hurwitz criterion and $a_1a_2 - a_0a_3 = 8(1 - b_0^2 - b_1 + b_0b_2)$. □

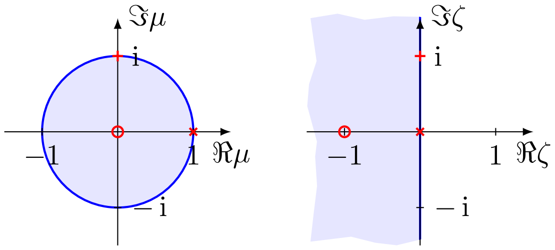


Figure 4.2: μ and ζ plane for the transformation in (4.21)

Lemma 4.14 (Strict stability (Erlicher et al., 2002, Arnold and Brüls, 2007))
The Newmark algorithm (4.2) is strictly stable at infinity provided that

$$\alpha_f < \frac{1}{2}, \quad \beta > \frac{\gamma}{2}, \quad \gamma > \frac{1}{2} \tag{4.22}$$

hold. Second order accurate methods are strictly stable at infinity if

$$\alpha_m < \alpha_f < \frac{1}{2}, \quad \beta > \frac{1}{4} + \frac{1}{2}(\alpha_f - \alpha_m).$$

Proof. The iteration matrix for $z \to \infty$ can be explicitly calculated

$$\mathbf{T}(\infty) := \lim_{z\to\infty} \mathbf{T}(z) = \begin{pmatrix} \frac{-\alpha_f}{1-\alpha_f} & 0 & 0 \\ \frac{-\gamma}{\beta(1-\alpha_f)} & 1-\frac{\gamma}{\beta} & 1-\frac{\gamma}{2\beta} \\ \frac{-1}{\beta(1-\alpha_f)} & -\frac{1}{\beta} & 1+\frac{1}{2\beta} \end{pmatrix}, \quad (\beta, 1-\alpha_f \neq 0). \tag{4.23}$$

Note that it does not depend on the parameter α_m. Clearly, one eigenvalue is given by the $(1,1)$-entry, which implies the stability condition $\alpha_f < \frac{1}{2}$. The conditions stated by Corollary 4.13 take the form

(i) $\dfrac{1}{\beta(1-\alpha_f)} > 0$,

(ii) $\dfrac{2(\gamma-\alpha_f)}{\beta(1-\alpha_f)} > 0$,

(iii) $\dfrac{4\beta - 1 - \alpha_f(4\gamma-2)}{\beta(1-\alpha_f)} > 0$,

(iv) $\dfrac{(2\alpha_f-1)(2\gamma-4\beta)}{\beta(1-\alpha_f)} > 0$,

(v) $\dfrac{(2\gamma-1)(\alpha_f - 2\alpha_f^2 - 2\beta + 2\alpha_f\gamma)}{4\beta^2(1-\alpha_f)^2} > 0$.

With $\alpha_f < \frac{1}{2}$, the first inequality may be simplified to (i') $\beta > 0$, which further reduces (ii) to (ii') $\gamma > \alpha_f$ and inequalities (iii-v) to (iii') $2\alpha_f + 4\beta - 4\alpha_f\gamma > 1$ and, as stated, $2\beta > \gamma$, $2\gamma > 1$. The conditions of (4.22) are enough since (i'-iii') are an upshot of the three inequalities. For the second set of inequalities we refer to the direct proof of Arnold and Brüls (2007, Lemma 1). □

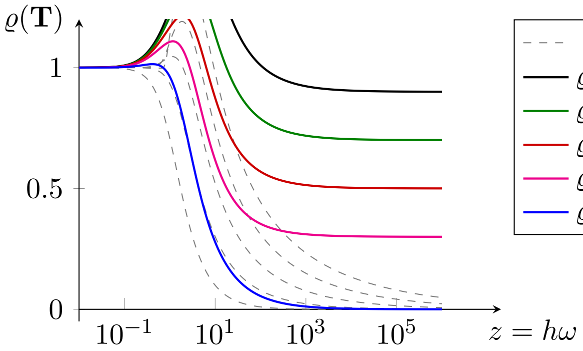


Figure 4.3: Newmark integrators that lack unconditional stability

Stability, even strict stability, at infinity and zero stability are nonetheless no sufficient conditions for (4.2) to be unconditionally stable. As an example, in Figure 4.3, we illustrate Newmark schemes that are stable at zero and infinity but possess an instability interval for certain values of $z = h\omega$, (I_d-stability, see (Schneider, 1995)). To construct the methods, we modified the parameter $\alpha_m \leq \frac{1}{2}$ of the CH(ϱ_∞) algorithm (4.4) while keeping all other parameters unchanged. Note that $\alpha_m < 0$ is no extraordinary parameter choice as for $\varrho_\infty < \frac{1}{2}$ also the CH(ϱ_∞) method uses such a setting and that the order of the above methods is just one.

Lemma 4.15 (Unconditional stability (compare Erlicher et al., 2002))
The Newmark algorithm (4.2) is unconditionally stable if it is stable at infinity and

$$\frac{1}{2} - \alpha_m + \alpha_f \leq \gamma \tag{4.24}$$

is fulfilled. For second order Newmark integrators, stability at infinity plus zero stability are equivalent to unconditional stability.

The proof employs the same techniques as the proof of Lemma 4.14 but the conditions are more involved since they depend on another value $z \neq 0$. Note that conditions (4.18) and (4.22) justify the assumptions β, $\gamma \neq 0$, α_m, $\alpha_f \neq 1$ respectively.

Definition 4.16.
Whenever the stability conditions (4.18), (4.22) and (4.24) are fulfilled, we will from now on simply refer to the integrators as *stable Newmark methods*.

Example 4.17 (Specific parameter choices and variants)

(a) *Trapezoidal rule and the ‘classical’ Newmark integrator:*
The only non-damping member of practical relevance is attained for the parameter choice $\beta = \frac{1}{4}$, $\gamma = \frac{1}{2}$, $\alpha_m = \alpha_f$, see Remark 4.1. Geometrically, this relates to the approximation of the state variables by means of using trapezoidals to approximate the underlying integral equations. In view of Lemma 4.10, the method is zero stable as well as unconditionally stable and so stable at infinity, but it lacks strict stability at infinity. The ‘classical Newmark’ or ‘Newmark-β integrator’ (Newmark, 1959, Erlicher et al., 2002) uses

$$\gamma = -\frac{1}{2} + \frac{2}{1+\varrho_\infty^2}\,, \quad \beta = \frac{(\gamma+1/2)^2}{4} = \frac{1}{(1+\varrho_\infty^2)}\,, \ \varrho_\infty \in [0,1]\,.$$

The conditions for stability at infinity and unconditional stability reduce to $\gamma \geq \frac{1}{2}$, $\beta \geq \frac{\gamma}{2}$, cf. Figure 4.4 (a).

(b) *The HHT and WBZ algorithms:*
The first algorithm in the present setting to combine second order accuracy and controllable damping is the so-called HHT(α_{HHT}) method of Hilber et al. (1977) which depends on one parameter $\alpha_{\mathrm{HHT}} \in [0, \frac{1}{3}]$.

$$\alpha_m = 0,\ \alpha_f = -\alpha_{\mathrm{HHT}}\,,\ \beta = \frac{(1-\alpha_{\mathrm{HHT}})^2}{4}\,,\ \gamma = \frac{1}{2} - \alpha_{\mathrm{HHT}}\,,$$

or, in terms of the numerical damping $\varrho_\infty \in [\frac{1}{2}, 1]$,

$$\alpha_m = 0,\ \alpha_f = \frac{1-\varrho_\infty}{1+\varrho_\infty},\ \beta = \frac{1}{(1+\varrho_\infty)^2},\ \gamma = \frac{3-\varrho_\infty}{2+2\varrho_\infty},\ \varrho_\infty \in [\tfrac{1}{2}, 1]\,.$$

Note that numerical damping ranges only from one to 0.5; instantaneous annihilation is not possible. To remedy this, Wood et al. (1981) proposed a family of algorithms (Bossak–Newmark, WBZ(α_{WBZ})) with $\alpha_f = 0$,

$$\alpha_m =: \alpha_{\mathrm{WBZ}} \in [-1, 0]\,,\ \alpha_f = 0\,,\ \beta = \frac{(1-\alpha_{\mathrm{WBZ}})^2}{4}\,,\ \gamma = \frac{1}{2} - \alpha_{\mathrm{WBZ}}\,,$$

or analogously

$$\alpha_m = \frac{\varrho_\infty - 1}{\varrho_\infty + 1}\,,\ \alpha_f = 0\,,\ \beta = \frac{1}{(1+\varrho_\infty)^2}\,,\ \gamma = \frac{3-\varrho_\infty}{2+2\varrho_\infty}\,,\ \varrho_\infty \in [0, 1]\,.$$

The parameters of the Bossak–Newmark method equal the ones of the Chung–Hulbert method for maximum damping $\varrho_\infty = 0$. For $z \to \infty$, the spurious root of the Bossak–Newmark algorithm always tends towards zero.

All the above algorithms are unconditionally stable and strictly stable at infinity if the trapezoidal rule $\varrho_\infty = 1$ is omitted. In Figure 4.4 (b), the parameter progressions for the classical schemes are illustrated. Note that, because for all methods (4.3) and $\beta = \frac{1}{4}(\frac{1}{2}+\gamma)^2$ are fulfilled, the plot is sufficient to describe the methods. The latter relation is called *optimal dissipation relation* in the literature and marks, where the principal roots become real for $z \to \infty$. This requirement has been the main design concept for the development of the three families. Note also that the trapezoidal rule is not unique in this representation: For the CH(ϱ_∞) method it is formally obtained with $\alpha_m = \alpha_f = \frac{1}{2}$, while for HHT and WBZ it corresponds to $\alpha_m = \alpha_f = 0$.

(c) *Third order consistent methods:*
As for applications in structural dynamics, simulation of constrained systems, and for the singularly perturbed mechanical systems from Chapter 3, it is unavoidable to rely on stable methods. The second Dahlquist barrier forbids to employ methods of higher order than two. Being dependent on four parameters, it is nevertheless possible to construct parameter settings such that third order consistency is attained. For methods from—or closely related to—(4.2), this has been done by Hilber and Hughes (1978) and Hoff and Pahl (1988a). Based on the representation of Kettmann (2009), the parameters may depend on

$$\alpha_f \in (-\infty, \tfrac{1}{6}(3-\sqrt{3})] \cup (\tfrac{1}{2}, \tfrac{1}{6}(3+\sqrt{3})]\,, \quad \alpha_m = \frac{12\alpha_f^2 - 6\alpha_f - 1}{12\alpha_f - 6}\,, \quad \beta = \frac{1}{6} + \frac{1}{2}(\alpha_f - \alpha_m)\,, \tag{4.25}$$

where the intervals for α_f are based on the zero stability condition (4.18).

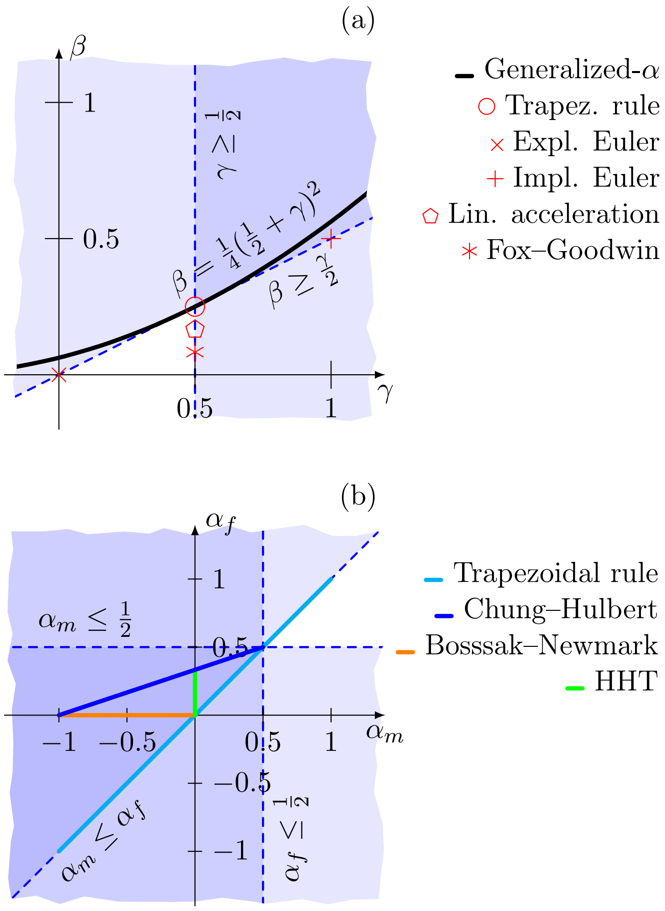


Figure 4.4: Parametric plots of parameters for Newmark, CH(ϱ_∞), HHT, and WBZ, illustrations adapted from (Chung and Hulbert, 1993, Géradin and Rixen, 2015)

In Figure 4.5 we illustrate that third order Newmark methods are not stable at infinity by inspecting the stability region using the test equation (4.14) and time step size $h = 1$. The shaded area is the domain of all pairs (ξ, ω) such that the according linear iteration possesses only stable solutions.

(d) *Explicit α-methods:*
The original proposal for the Newmark integrators (as classical Newmark scheme, HHT method or WBZ/Bossak–Newmark) was only formally given in the fully implicit form as it is presented here. Due to restricted computer capabilities in the early days of computational structural dynamics, an explicit implementation of the algorithms was favored. Later on, this approach was occasionally still considered because for certain applications like wave-propagation or impact problems the highly nonlinear character of the equations requires very small step sizes anyway or because of memory restrictions for the linear algebra overhead in large-scale simulations (see for example the *CDTire* model of Gallrein et al. (2014)). For the proposed family of Newmark integrators, alternative parameter settings have been developed by Hulbert and Chung (1996) where the aim was to obtain a maximum stability interval.

Another reason to study explicit methods is that for splitting techniques it is opportune to use the same base algorithm for the different integrators used. Hughes and Liu (1978)

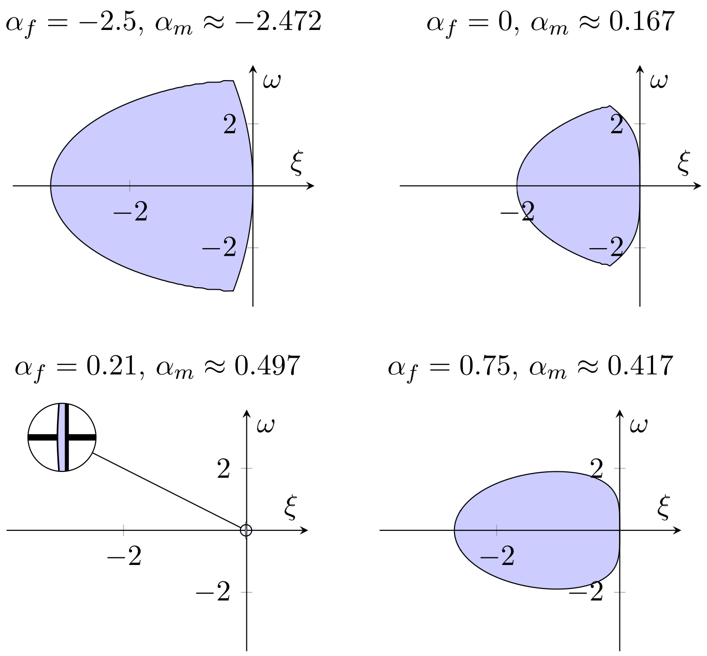


Figure 4.5: Stability regions for third order consistent Newmark methods according to (4.25)

optimized the parameters of the HHT method for a general implicit-explicit splitting algorithm. For the Newmark algorithms in the present form, these results were generalized by Daniel (2003).

(e) *Generalized-α for first order systems:*
The basic idea of (4.2) to use a collocation condition at a shifted time instance by imposing equality of the weighted sums of 'correct' derivatives and 'derivative-like' variables was used by Jansen et al. (2000) to construct a family of methods for first order systems that also allows for controllable numerical damping. In the context of general mechatronic systems, cf. (2.16), where first and second order methods are coupled, the algorithm has been applied by Brüls and Arnold (2008). For the system

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\chi}(\boldsymbol{x}(t)) \quad \text{resp.} \quad \dot{\boldsymbol{x}}(t) = \boldsymbol{\chi}(\boldsymbol{x}(t), \boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)) \,,$$

we introduce 'derivative-like' variables $\boldsymbol{a}_n^{\boldsymbol{x}} \in \mathbb{R}^{n_{\boldsymbol{x}}}$, $n = 0, 1, \ldots$, and the θ-method update

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + h(1 - \gamma^{\boldsymbol{x}})\boldsymbol{a}_n^{\boldsymbol{x}} + h\gamma^{\boldsymbol{x}}\boldsymbol{a}_{n+1}^{\boldsymbol{x}} \,,$$

where $\boldsymbol{a}_n^{\boldsymbol{x}}$ and the actual derivative variables $\dot{\boldsymbol{x}}_n := \boldsymbol{\chi}|_{\boldsymbol{x}=\boldsymbol{x}_n}$ are related through

$$(1 - \alpha_m^{\boldsymbol{x}})\boldsymbol{a}_{n+1}^{\boldsymbol{x}} + \alpha_m^{\boldsymbol{x}}\boldsymbol{a}_n^{\boldsymbol{x}} = (1 - \alpha_f^{\boldsymbol{x}})\dot{\boldsymbol{x}}_{n+1} + \alpha_f^{\boldsymbol{x}}\dot{\boldsymbol{x}}_n \,.$$

The 'optimal' parameter setting in the sense of Section 4.2.2 below is

$$\alpha_m^{\boldsymbol{x}} := \frac{1}{2}\frac{3\varrho_\infty - 1}{\varrho_\infty + 1} \,, \quad \alpha_f^{\boldsymbol{x}} := \alpha_f = \frac{\varrho_\infty}{\varrho_\infty + 1} \,,$$

and the second order condition $\gamma^{\boldsymbol{x}} = \frac{1}{2} - \alpha_m^{\boldsymbol{x}} + \alpha_f^{\boldsymbol{x}}$ remains unchanged. Note that for $\varrho_\infty = 1$, the algorithm again coincides with the trapezoidal rule and that for maximum dissipation $\varrho_\infty = 0$, the method is BDF(2), cf. (2.28).

$\diamondsuit$

4.2.2 Optimal parameter choice

Remark 4.18 (Damped oscillators)
For phase error analysis and certain applications using linear stability investigations, sometimes the problem class of damped oscillators with harmonic excitation

$$\ddot{q}(t) = -2\xi\omega\dot{q}(t) - \omega^2 q(t) + \sin(\Omega t)\,, \quad \xi, t, \Omega \geq 0,\, \omega > 0\,, \tag{4.26}$$

is used to study the behavior of numerical integrators (Arnold et al., 2011). Especially linear resonance effects in vehicle system dynamics are captured by this approach. The parameter ξ is apparently used to control the amount of physical damping in the system; $\xi = 0$ corresponds to the undamped case from (4.12) *while $0 < \xi < 1$ is called the* underdamped *and $\xi > 1$ the* overdamped case. *The* critical damping *at $\xi = 1$ plays a special role since here the eigensystem of the dynamics degenerate (Hughes, 1987). Even though this test equation is more complex and therefore captures more effects and possible issues for real-world problems, it should be pointed out that for general linear systems* (4.7)*, a transformation to a one-degree-of-freedom system is only possible if a common set of eigenvectors can be found. Note that some researchers from structural and molecular dynamics construct and optimize their methods only for undamped systems or use certain heuristics to approximate the load vector such that the order may drop in the presence of dissipative terms $\xi > 0$ or complex external excitations, see for example (Bazzi and Anderheggen, 1982).*

The starting point to determine suitable parameters for the Newmark integrators is the scalar damped oscillator equation (4.26) without external force:

$$\ddot{q}(t) + 2\xi\omega\dot{q}(t) + \omega^2 q(t) = 0\,, \quad q(0) = q_0\,,\ \dot{q}(0) = \dot{q}_0 \tag{4.27}$$

with the analytic solution

$$q(t) = \mathrm{e}^{-\xi\omega t}\left(C_1\cos(\bar{\omega}t) + C_2\sin(\bar{\omega}t)\right)\,, \quad C_1 = q_0\,,\ C_2 = (q_0\xi\omega + \dot{q}_0)/\bar{\omega}\,, \quad \bar{\omega} = \omega\sqrt{1-\xi^2} \tag{4.28}$$

in case $0 \leq \xi < 1$ and

$$q(t) = \mathrm{e}^{-\xi\omega t}\left(C_1\exp^{\hat{\omega}t} + C_2\exp^{-\hat{\omega}t}\right)\,, \quad C_{1|2} = \frac{q_0(\hat{\omega} \pm \xi\omega) \pm \dot{q}_0}{2\hat{\omega}}\,, \quad \hat{\omega} = \omega\sqrt{\xi^2 - 1}$$

for strong damping $\xi > 1$.

Either way, for positive values of ξ, the analytic solution of (4.27) decays exponentially fast in time. Note, however, that the damping properties of the numerical solution always resemble the case of infinite stiffness for the undamped case (4.15), i. e., for $h\xi \to \infty$ the onestep amplification is as large as the numerical damping for $h\omega \to \infty$, where from a geometric point of view annihilation should occur. If no infinite frequency but the damping forces, i. e., the energy dissipating parts are considered to have the most important impact on the system, the parameter $\xi > 1$ in (4.26) becomes dominant. So, in this case the analysis may instead be carried out using the test equation $\ddot{q}^\delta(t) + \delta^{-1}\dot{q}^\delta(t) = 0$. We postpone the (easier) analysis of this case to Example 5.27 below.

The discrete analogue of (4.28) in the case $0 \leq \xi < 1$ can be obtained if the onestep mapping from $(q_n, hv_n, h^2a_n)^\top$ to $(q_{n+1}, hv_{n+1}, h^2a_{n+1})^\top$ of (4.2) for (4.27) is transformed to a difference equation for the position variables q_n. We assume that its characteristic equation has two complex conjugate principal roots $\mu_{1|2} = \mathrm{e}^{\tilde{\Omega}(-\tilde{\xi}\pm\mathrm{i})}$ which uniquely define $\tilde{\Omega} > 0$ and $\tilde{\xi} \in \mathbb{R}$.

Denoting the one spurious root by μ_3, there exist constants $\tilde{C}_i \in \mathbb{C}$, $i = 1, 2, 3$, such that q_n may be expressed as

$$q_n = \mathrm{e}^{-\tilde{\xi}\tilde{\omega}t_n} \cdot (\tilde{C}_1 \cos(\tilde{\omega}t_n) + \tilde{C}_2 \sin(\tilde{\omega}t_n)) + \tilde{C}_3\mu_3^n\,, \quad \tilde{\omega} := \tilde{\Omega}/h\,. \tag{4.29}$$

A comparison with (4.28) shows that—if the influence of $\tilde{C}_3\mu_3^n$ is negligible—the accuracy may be expressed in terms of

$$P := \frac{\bar{\omega}}{\tilde{\omega}} - 1 \quad \text{(numerical dispersion)}$$

$$\text{and} \qquad \Delta\xi := \tilde{\xi} - \xi \quad \text{(algorithmic damping error, numerical dissipation).}$$

The value P is also called *relative period error* because for the periods $\bar{T} := 2\pi/\bar{\omega}$ and $\tilde{T} := 2\pi/\tilde{\omega}$ it holds $P = (\tilde{T} - \bar{T})/\bar{T}$. Sometimes, the relative period error is used to implement heuristics for local error control in adaptive step size implementations (Cardona and Géradin, 1994). The idea of using dispersion and dissipation for accuracy judgment goes back to the work of Bathe and Wilson (1973), where the authors studied period elongation and numerical damping for certain parameter settings in the investigation of Newmark-type and related methods. From the comparison of (4.28) and (4.29), it is evident why optimally zero stable methods are often favored: The term $\tilde{C}_3\mu_3^n$ in this case is, in fact, no longer present. Note that the notions of dissipation and dispersion only make sense for $\xi < 1$.

Remark 4.19 (Design paradigms (Hughes, 1987))
According to the given reference, an ODE time integration method for structural mechanics should serve the following properties:

(a) *It should be (at least) second order accurate.*

(b) *It should show unconditional linear stability.*

(c) *Each time integration step should neccesitate the solution of no more than one set of implicit equations of dimension $n_{\boldsymbol{q}}$.*

(d) *It should be self-starting, i. e., except of the initialization of all involved quantities from the initial data, no further computations should be necessary to start the time integration.*

(e) *The algorithmic damping of high frequency modes should be controllable by the user using one parameter.*

Of course, there are many adaptations to this collection and, depending on the application, researchers value some of these properties higher than others. Requirement (a) has already been discussed in Section 2.2.1: Due to modest smoothness of input values, event handling, and the typical accuracy requirement, second order methods are mostly preferred. In view of the second Dahlquist barrier, requisite (b) even accentuates this fact. For integrators (4.2)*, the analysis is stated in Lemma 4.15. In a nutshell, the goals of requirements (c) and (d) are on the one hand to keep the computational effort low, on the other hand to exclude algorithms that demand immense changes in very involved existing codes. We will see in Section 6.1 that the linear structure of* (4.2) *ascertains requirement (c). Self-starting is important since an extension to variable step size schemes should be possible in a straightforward manner. Variable time step size implementations for multistep methods are very complex such that onestep algorithms are commonly favored or, as Brüls (2005) puts it: 'Analysts from structural dynamics are sometimes reluctant with respect to this approach [BDF methods], fearing the computational burden of a multi-stage algorithm.' We have already pointed out that* (4.2) *is a multistep methods because* *the acceleration-like variables $\boldsymbol{a}_n$ enter the equations. Nevertheless, from a computational point of view, the methods are easily implemented as if they were onestep methods; depending on the initialization of $\boldsymbol{a}_0$ they fulfill (d) but variable step size implementations are yet non-trivial. There are several proposals in the literature: Negrut et al. (2005) propose to adapt $\boldsymbol{a}_n$ in each time step based on the step size change for HHT, an approach that was later extended to Newmark integration methods of more general form by Jay and Negrut (2008). Brüls and Arnold (2008) suggest to alter the parameter γ in each step while Rang (2013) chooses γ and α_m variable.*

If (4.3), (4.18), (4.22), *and* (4.24) *are fulfilled, the Newmark integrators meet (a)–(d) such that for an explanation of the parameters from CH(ϱ_∞), HHT, and WBZ we will in the following concentrate on requirement (e).*

According to Hilber et al. (1977), requirement (e) in particular means that it should be possible to attain parameter values such that the algorithm is non-damping and then may be smoothly varied towards more and more damping of high frequencies. At best, it should even be possible to obtain an L-stable method if needed as this is better for dissipation-dominated problems. For the case of low values of $z = h\omega$, dispersion and damping error become more important. Here, the amplification factor should remain close to one. Smoothness is also important with respect to the frequency range: As z varies, there should be no "cusps in the numerical damping" (Hoff and Pahl, 1988a, Chung and Hulbert, 1993). As a result, for all finite values of z

(i) there should be two principal roots as bifurcation points are typically nonsmooth and

(ii) the spurious root should remain smaller in absolute value than the principal roots.

After inserting the second order condition (4.3), the characteristic polynomial of $\mathbf{T}(\infty)$, see (4.16) and (4.23), can be written as

$$\chi_{\mathbf{T}(\infty)}(\mu) = -\mu^3 + \left(3 + \frac{1}{\alpha_f - 1} + \frac{\Delta_\alpha - 1}{\beta}\right)\mu^2 + \left(\frac{\alpha_m + \alpha_f(2\Delta_\alpha + 3\beta) + \beta}{\beta(\alpha_f - 1)}\right)\mu + \frac{\alpha_f(\beta + \Delta_\alpha)}{\beta(\alpha_f - 1)}\,, \tag{4.30}$$

with two principal roots $\mu_{1|2}^\infty$ whose imaginary parts vanish iff the *optimal dissipation relation* $\beta = \frac{1}{4}(\frac{1}{2} + \gamma)^2$ is fulfilled, cf. Example 4.17 (b). In that case we have

$$\lim_{z\to\infty} \mu_{1|2} =: \mu_1^\infty = \mu_2^\infty = \frac{\alpha_f - \alpha_m - 1}{\alpha_f - \alpha_m + 1}\,, \quad \lim_{z\to\infty} \mu_3 =: \mu_3^\infty = \frac{\alpha_f}{\alpha_f - 1} \text{ as spurious root.}$$

For the HHT and WBZ method, either the parameter α_m or the parameter α_f is not present, such that the algorithm is already fully characterized by the requirement $|\mu_1^\infty| = \varrho_\infty$. There are actually two possibilities to choose the parameters and fulfill this relation. $\mu_1^\infty = -\varrho_\infty$ is the one with better low-frequency behavior which is evident since only then for $\varrho_\infty = 1$ the trapezoidal rule is obtained. As a result, the HHT parameters may be seen as optimal for all methods (4.2) with $\alpha_m = 0$ which implies a vanishing spurious root for $z \to 0$. Still, it is not possible to additionally control the spurious root for large frequencies such that the range of numerical damping for HHT is limited to $\varrho_\infty \in [\frac{1}{2}, 1]$. For WBZ, this drawback is resolved, yet the low-frequency behavior is slightly worse.

With the introduction of an additional parameter, Chung and Hulbert (1993) had the possibility to compromise between these two objectives. They also chose α_m such that the principal roots coincide for $z \to \infty$ at $-\varrho_\infty$ and found that dissipation for moderate z is minimized for $\alpha_f = \frac{\alpha_m+1}{3}$, which corresponds to $\mu_3^\infty = -\varrho_\infty$. This leads to the choices of (4.3) and (4.4). In Figure 4.6 the numerical dissipation, dispersion, and damping for the three classical settings is compared. For all three categories, the CH(ϱ_∞) setting performs best.

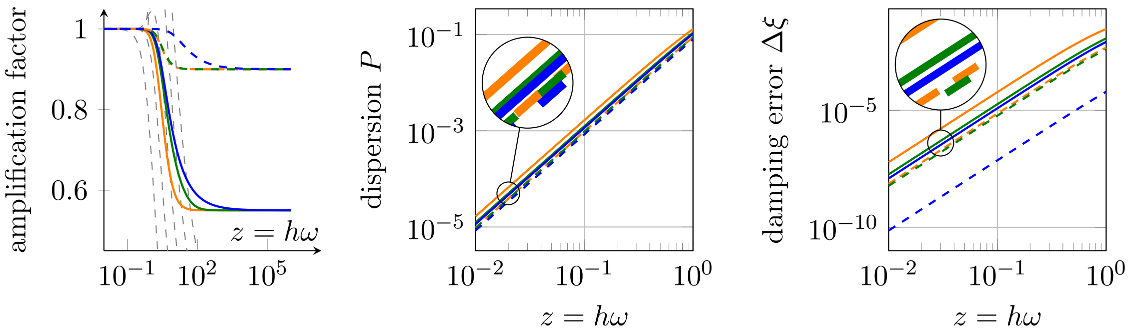


Figure 4.6: Damping, dispersion and damping error of CH(ϱ_∞) (blue), HHT (green), WBZ (orange) for parameter $\varrho_\infty = 0.55$ (solid) and $\varrho_\infty = 0.9$ (dashed) being applied to $\ddot{q} + \omega^2 q = 0$

Remark 4.20 (Benefits of controllable damping)
There is an ongoing debate whether or not numerical damping is a favorable property of an algorithm. The discussion is heavily influenced by the strong development in structure preserving integrators in the last decade. In fact, numerical damping corresponds to an artificial energy decay in the linear system (4.7) *and Newmark (1959) himself argued that for 'his' algorithm* (4.2a), (4.2b) *parameters should be chosen in a way that 'unnatural' damping effects are minimized. One might also argue that it is always possible to simply add physical, so-called 'viscous' damping to the model such that the behavior of the analytic solution becomes more stable. This approach, nevertheless, has two disadvantages: First, it is unclear how to choose a correct physical damping model and the involved parameters, and, secondly, instabilities and/or resonances in the original model may remain undetected. Concerning the statement that smooth solutions appear physically more correct and are needed for further computations, one can add filtering techniques as a postprocessing task. Finally, one should also take the influence of numerical damping on other aspects of the overall simulation into account, as 'numerical damping and step size control are in some sense contradictory goals' (Simeon, 1998).*

Remark 4.21 (Overshoot-phenomenon)
The seemingly optimal parameters of the CH(ϱ_∞) method suffer from a phenomenon that is typically observed for problems with large eigenfrequencies and simulations with relatively large time steps. Overshoot *describes the tendency of a method to strongly overestimate the response of the mechanical system in the first integration steps, mostly indicated by spurious oscillations. As a result, the order of the method may drop in a transient phase while afterwards it is numerically preserved. Sometimes, the impact of the initial artifacts is so severe that even a loss of second order accuracy is taken (Sanborn et al., 2014). For a long time, overshoot has been mostly excluded from theoretical investigations and researchers used only numerical benchmarks to observe whether overshoot is an issue (Hoff and Pahl, 1988b). For these experiments, energy-measures of the numerical solution are monitored. The first ones to systematically analyze overshoot were Hilber and Hughes (1978) who argued that the energy growth in the transient phase is due to large elements in powers of the amplification matrix* $\mathbf{T}(z)$ *for* $z \to \infty$*. Even when the spectral radius is small, powers of* $\mathbf{T}(\infty)$ *and its transformation matrix to canonical form may become very large. Hoff and Pahl (1988a) also explained overshoot by the high condition number of the transformation matrix. Overshoot for the present version of Newmark integrators has been extensively studied by Erlicher et al. (2002).*

The first to explain overshoot by the degenerate Jordan-structure of $\mathbf{T}(\infty)$ *were Cardona and Géradin (1989) in their analysis of HHT for constrained mechanical systems. The Jordan* *canonical forms*

$$\mathbf{T}(\infty) = \mathbf{C}\mathbf{J}_*\mathbf{C}^{-1} \tag{4.31}$$

of the CH(ϱ_∞), HHT, and WBZ algorithm are given by

$$\begin{aligned}
\mathbf{J}_{\mathrm{CH}} &= \begin{pmatrix} -\varrho_\infty & 1 & 0 \\ 0 & -\varrho_\infty & 1 \\ 0 & 0 & -\varrho_\infty \end{pmatrix}, \\
\mathbf{J}_{\mathrm{HHT}} &= \begin{pmatrix} -\varrho_\infty & 1 & 0 \\ 0 & -\varrho_\infty & 0 \\ 0 & 0 & (\varrho_\infty - 1)/(2\varrho_\infty) \end{pmatrix}, \quad \left(\varrho_\infty > \frac{1}{2}\right), \\
\mathbf{J}_{\mathrm{WBZ}} &= \begin{pmatrix} -\varrho_\infty & 1 & 0 \\ 0 & -\varrho_\infty & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\end{aligned} \tag{4.32}$$

such that for the matrix powers $(\mathbf{T}(\infty))^n$ the non-diagonal entries cause an amplification as the values $(\mathbf{J}^n_{\mathrm{CH}})_{1,3} = \frac{1}{2}n(n-1)(-\varrho_\infty)^{n-2}$, $(\mathbf{J}^n_{\mathrm{HHT/WBZ}})_{1,2} = n(-\varrho_\infty)^{n-1}$ may grow large unless the damping parameter is chosen very small, cf. Figure 4.7 below. This problem is fairly generic: Whenever the second order- and the optimal dissipation relation are fulfilled, the principal roots have a degenerate eigenspace structure for $z \to \infty$ such that overshoot may be observed. This also holds true for the first order classical Newmark method, where the canonical form of the amplification matrix consists of one Jordan block $\mathbf{J}_{\mathrm{Newm}} \in \mathbb{R}^{2\times 2}$. Note, however, that the long term behavior of the integrators is not influenced by the overshoot phenomenon.

A new parameter set with optimized overshoot behavior *It is possible to adapt the parameter choice of Chung and Hulbert (1993) such that three Jordan blocks are present while keeping the absolute values of the eigenvalues in the limit case and the low frequency behavior is only slightly influenced: To this means, we drop the requirement that all eigenvalues of the linear recursion in* (4.16) *coincide for the limit of infinite stiffness $z \to \infty$ but instead just place them on the complex unit circle of radius ϱ_∞. Guided by Dahlquist's result, it seems reasonable that for $\varrho_\infty = 1$ still the trapezoidal rule is attained. So, we add another parameter $\phi_0 \in [0, \pi]$ and the requirement*

$$\arg(\lim_{z\to\infty} \mu_1(\mathbf{T}(z))) \overset{!}{=} \pi - (\pi - \phi_0)(1 - \varrho_\infty), \quad \varrho_\infty \in (0, 1].$$

Keeping $\alpha_f = \varrho_\infty/(\varrho_\infty + 1)$, such that the spurious root tends towards $-\varrho_\infty$ for $z \to \infty$, we attain a nonlinear system for the parameters with the solution

$$\left.\begin{aligned}
\alpha_m &= \frac{\varrho_\infty}{\varrho_\infty + 1} + \frac{\varrho_\infty^2 - 1}{1 + \varrho_\infty^2 - 2\varrho_\infty \cos(\phi_0 - \phi_0\varrho_\infty + \pi\varrho_\infty)}, \\
\alpha_f &= \frac{\varrho_\infty}{\varrho_\infty + 1}, \\
\beta &= \frac{1}{1 + \varrho_\infty^2 - 2\varrho_\infty \cos(\phi_0 - \phi_0\varrho_\infty + \pi\varrho_\infty)}, \\
\gamma &= \frac{1}{2} - \alpha_m + \alpha_f.
\end{aligned}\right\} \mathit{Gen}(\varrho_\infty, \phi_0)$$

For $\phi_0 = \pi$, the methods are identical to the CH(ϱ_∞) setting. The numerical dissipation and dispersion behavior of these methods is slightly inferior compared to CH(ϱ_∞) but the overshoot in the sense of the above growth of Jordan block powers is diminished. The growth of powers of

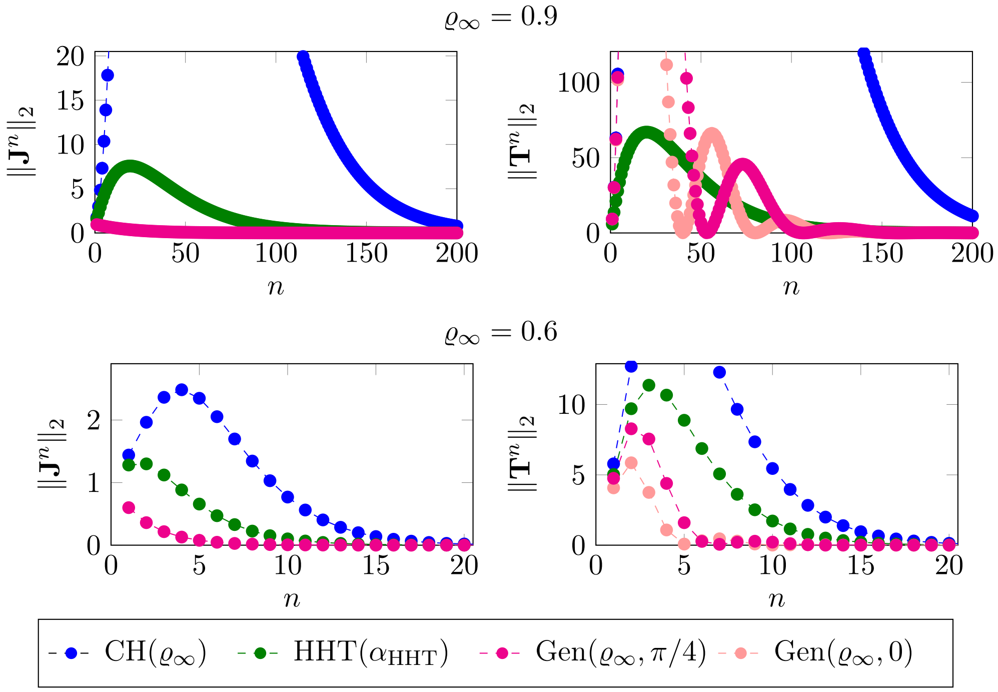


Figure 4.7: Norm of powers of the Jordan matrices for CH(ϱ_∞), HHT and Gen(ϱ_∞, ϕ_0)

the Jordan canonical forms for CH(ϱ_∞), HHT(α_{HHT}) and Gen(ϱ_∞, ϕ_0) is shown in Figure 4.7. The off-diagonal entries clearly reveal a growth in the first steps of the iteration. Note that for Gen(ϱ_∞, ϕ_0) the plot is independent of the parameter ϕ_0.

It should be pointed out that only the growth in matrix powers of $\mathbf{J}$ *is not sufficient to completely characterize overshoot. The numerical experiments in Chapter 6 will indeed show that overshoot is not as much prevented for the 'improved' parameter set of* Gen(ϱ_∞, ϕ_0) *as the left plots would have indicated. The right plots in Figure 4.7 illustrate the growth of powers of the amplification matrices* $\mathbf{T}(\infty)$ *themselves rather than just their canonical forms and show that still there is an amplification even for single eigenvalues. This comes without surprise since for $\phi_0 \to \pi$ the algorithm is identical to the CH(ϱ_∞) parameter set with its large degenerate Jordan structure such that the condition number of the transformation matrix to Jordan canonical form is very large (see the example of Golub and Van Loan, 1996, Sect. 7.1.5).*

In the literature, position-, velocity-, acceleration-, and energy-overshoot are sometimes distinguished and it is well-known that nonlinear effects and the particular problem formulation influence spurious oscillations. Briefly, the errors from the initialization phase *of the method are probably even more important than the the Jordan-structure. Note that the consistent initialization of mechanical systems is a non-trivial task and that the singular perturbation terms for the problem classes of Chapter 3 also make a 'stable initialization' challenging.*

In Figure 4.8 the stability regions in the sense of Gladwell and Thomas (1980) for the four settings are compared. HHT($\alpha_{\text{HHT}} = -\frac{1}{7}$) and WBZ($\alpha_{WBZ} = -\frac{1}{7}$) both have numerical damping $\varrho_\infty = \frac{3}{4}$ for $z \to \infty$.

Remark 4.22 (Algorithms with optimized damping behavior)

(a) Blended Lobatto methods
Schaub and Simeon (2003) propose to use a class of super partitioned additive Runge–Kutta methods called SPARK (Jay, 1999) to transfer the idea of adjustable algorithmic

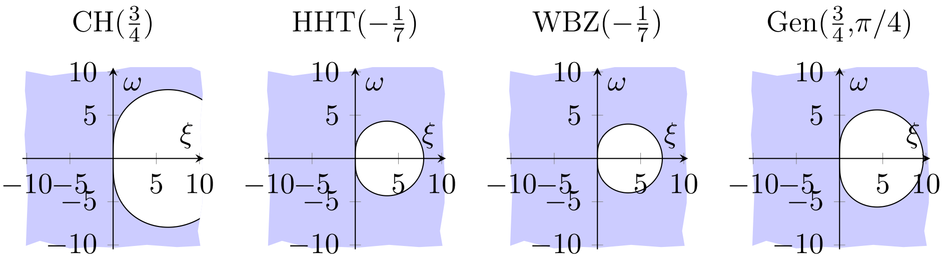


Figure 4.8: Stability regions of Newmark integrators

damping to a class of Runge–Kutta methods. The basic idea of this blended approach lies in using convex combinations of Runge–Kutta type methods with different stability properties: Given $k > 1$ Runge–Kutta methods, e. g. of Lobatto-III-type (Hairer and Wanner, 2002), with parameters $(a_{ij}^{(k)}, c_j, b_i)$, $k = 1, \dots, n_{\mathrm{RK}}$, $i, j = 1, \dots, s$, with the same weight vector $(b_i)_{i=1,\dots,s}$ and knot vector $(c_j)_{j=1,\dots,s}$ it can be shown that under very weak conditions the convex combined method, i. e., the one with parameters

$$(a_{ij}^{\mathrm{bl}}, c_j, b_i) \quad \textit{with} \quad a_{ij}^{\mathrm{bl}} := \sum_{k=1}^{n_{\mathrm{RK}}} \nu_k a_{ij}^{(k)}, \quad \left(\sum_{k=1}^{n_{\mathrm{RK}}} \nu_k = 1,\ \nu_k \geq 0,\ k = 1, \dots, n_{\mathrm{RK}} \right) \tag{4.33}$$

inherits (stage-) order along with other advantageous properties from the underlying methods. The proof is based on the concept of simplifying assumptions introduced by Butcher (Hairer and Wanner, 2002, Chap. IV). Lobatto-IIIC methods are predetermined for the use in the SPP case since they are stiffly accurate methods (Prothero and Robinson, 1974) which usually outperform other Runge–Kutta methods for very stiff equations or high-index DAE systems.

Damping controlled blended Lobatto *It is important to notice that within the framework of blended Lobatto methods it is not only possible to adjust the numerical damping of the algorithms at high frequencies, but also to control the frequency range of low damping. In Figure 4.9 (a) we depicted the numerical amplification for six examples with $s = 2$ stages and the linear test equation (4.12). For Runge–Kutta methods this is just the absolute value of the stability function. For a blending of two-stage Lobatto-IIIC and IIID methods with predefined numerical damping factor ϱ_∞ for $z \to \infty$, denoted by C-D(ϱ_∞), one can use $\nu_{\mathrm{C}} = \frac{1-\varrho_\infty}{1+\varrho_\infty}$ in (4.33) but may obtain non-monotone dissipation behavior along the imaginary axis. More details on this approach can be found in (Schaub, 2004, Simeon, 2013).*

For a blend of A, C, and D methods and $s = 2$, one can also set the value z_0 of $h\omega$ where the amplification factor is exactly $\frac{1}{2}(1+\varrho_\infty)$ (denoted by A-C-D(ϱ_∞, z_0)) using the weights

$$\begin{aligned} \nu_{\mathrm{A}} &= \frac{\eta_1 z_0^2 + 4\varrho_\infty \eta_2 - \sqrt{(\varrho_\infty + 1)^2 \eta_2 (2\varrho_\infty + z_0^2 + 6)(\eta_1 - 2(\varrho_\infty - 1)^2)}}{\eta_1 z_0^2 - (\varrho_\infty - 1)^2 \eta_2}, \\ \nu_{\mathrm{D}} &= -\frac{2\varrho_\infty \nu_{\mathrm{A}} - \varrho_\infty}{\varrho_\infty + 1}, \end{aligned}$$

where

$$\eta_1 := (3\varrho_\infty + 1) z_0^2, \quad \eta_2 := (\varrho_\infty + 3) z_0^2. \tag{4.34}$$

Note that having an enlarged low-damping region comes at the cost of the method being more and more like Lobatto-IIIA, i. e., with singular Runge–Kutta matrix, cf. Remark 3.20. For $s = 2$, Lobatto-IIIA is simply the trapezoidal rule. Using a blending that contains Lobatto-IIID results in a drop of the stage order of the methods.

(b) ***An SDIRK method with controllable damping***
The blended Runge–Kutta methods nevertheless bear the drawback that for most families of practical interest, as the Radau and Lobatto methods, the Runge–Kutta matrices do not show a structure allowing for drastical savings of computational effort in terms of the linear algebra overhead. So, in each time integration step rather large nonlinear systems need to be solved. To remedy this, one might consider using a $\underline{\text{s}}$*ingly* $\underline{\text{d}}$*iagonally* $\underline{\text{i}}$*mplicit* $\underline{\text{R}}$unge–$\underline{\text{K}}$utta *type method (SDIRK) and forsake the concept of blended methods. Since for $s = 1$, the only second order Runge–Kutta method is the non-damping midpoint rule, we will consider a two stage method with Butcher tableau*

$$\begin{array}{c|cc} c_1 & \gamma_{\text{RK}} & \\ c_2 & a_{21} & \gamma_{\text{RK}} \\ \hline & b_1 & b_2 \end{array} .$$

Guided by the design paradigms for Newmark methods in Remark 4.19, yet not strictly sticking to it, the method should be (exactly) second order accurate and A-stable and the absolute value of the stability function at infinity should equal a user-defined value ϱ_∞. Being a onestep method, self-starting is no issue but in view of Remark 3.20, we have to mention that more than stage order one cannot be obtained for these methods. Before explaining the construction of the method in detail, we emphasize that it will mainly be used to have a fair comparison with a onestep method when studying the numerical properties of Newmark integrators in Chapter 6 below. Following the 'address accuracy before stability'-rule (Hulbert and Chung, 1996), at first we have the two conditions for second order

$$\gamma_{\text{RK}} = \frac{1}{2} - a_{21} b_2 \,, \quad b_1 = 1 - b_2 \,.$$

The limit of the stability function then is

$$\lim_{z\to\infty} R(z) = \frac{4a_{21}b_2(a_{21}b_2+1)-1}{(1-2a_{21}b_2)^2} \overset{!}{=} \pm\varrho_\infty$$

and necessarily real. So, to have controllable damping, we have to solve this system, resulting in four solution branches, e. g. for a_{21} such that only b_2 is left as a free parameter. As we are only interested in second order methods (third order, even for linear systems, is not attainable anyway), it seems reasonable to choose b_2 such that the error constant for the nonlinear third order error term, i. e., the one corresponding to elementary differential with tree *(Hairer et al., 1993), is minimized leading once again to four solution branches only depending on ϱ_∞, three of which would lead to unbounded parameters or unbounded linear error constants (corresponding to* *). The method's parameters for the fourth solution are explicitly given by*

$$\gamma_{\text{RK}} = \frac{1}{(2+\sqrt{2}\sqrt{1+\varrho_\infty})}\,, \qquad a_{21} = \frac{1-2\sqrt{2}\sqrt{1+\varrho_\infty}+\varrho_\infty(3+2\varrho_\infty-\sqrt{2}\sqrt{1+\varrho_\infty})}{3(\varrho_\infty^2-1)}\,,$$

$$b_2 = \frac{3}{2(2+\frac{1}{1+\varrho_\infty}+\frac{\sqrt{2}}{\sqrt{1+\varrho_\infty}})}\,, \quad b_1 = 1 - b_2 \,.$$

It is easily verified that for all values of $\varrho_\infty \in [0,1]$, this method has bounded parameters, even all in $(0,1)$, that the Runge–Kutta matrix is invertible (its determinant is γ_{RK}^2), and that the corresponding methods are A-stable. The latter assertion can be seen from the fact that all poles of the Runge–Kutta matrix are in the right complex halfplane and the method is I-stable (Hairer and Wanner, 2002). As $\gamma_{\mathrm{RK}} \in \mathbb{R}_{>0}$, no eigenvalues of the Runge–Kutta matrix lie on the imaginary axis as well. The numerical damping behavior for $\varrho_\infty \in \{0.3,\, 0.8\}$ of the proposed SDIRK methods is shown in Figure 4.9 (b). Similar, but always L-stable, methods are proposed and analyzed by Owren and Simonsen (1995).

(c) The BDF-α method

Celaya and Anza (2013) motivate to extend the design ideas of the HHT scheme to generalize the backward differentiation formulae (BDF) to a time integration method with controllable damping. Like for the design of Newmark integrators (4.2) *the key idea is a weighted-sum approach between trapezoidal rule and an L-stable scheme (here: BDF(2)). The methods fall into the broader context of A-BDF methods as they are proposed by Fredebeul (1998). Depending on a parameter α_{BDF} as for HHT, the method is defined as*

$$(\tfrac{3}{2}+\alpha_{\mathrm{BDF}})\boldsymbol{x}_{n+2} - 2(1+\alpha_{\mathrm{BDF}})\boldsymbol{x}_{n+1} + (\tfrac{1}{2}+\alpha_{\mathrm{BDF}})\boldsymbol{x}_n = h\left((1+\alpha_{\mathrm{BDF}})\boldsymbol{\chi}_{n+2} - \alpha_{\mathrm{BDF}}\boldsymbol{\chi}_{n+1}\right)\,,$$

where $\alpha_{\mathrm{BDF}} := -\varrho_\infty/(1+\varrho_\infty) \in [-0.5,\, 0]$ depends on the user-defined damping ratio ϱ_∞ and we considered the first order system (2.4). *Note that the generalized-α method for first order ODEs as proposed by Jansen et al. (2000) also defines a smooth transition from midpoint-rule to BDF(2).*

(d) Super implicit BDF

Vater et al. (2011) also introduced alterations on BDF schemes in the context of wave phenomena. The approach may again be seen as a blending of different time integration schemes or an extension of BDF-α. Since a direct application of these integration schemes, in connection with so-called super implicit BDF/extreme BDF/replica algorithms, does not yield zero stable algorithms, we will not discuss them any further here but simply state that in the context of partial differential equations the notions of numerical damping, dissipation, and optimized dispersion are still today a field of vivid research. (The lack of zero stability is compensated by switching to the trapezoidal rule in dependence of local CFL numbers.)

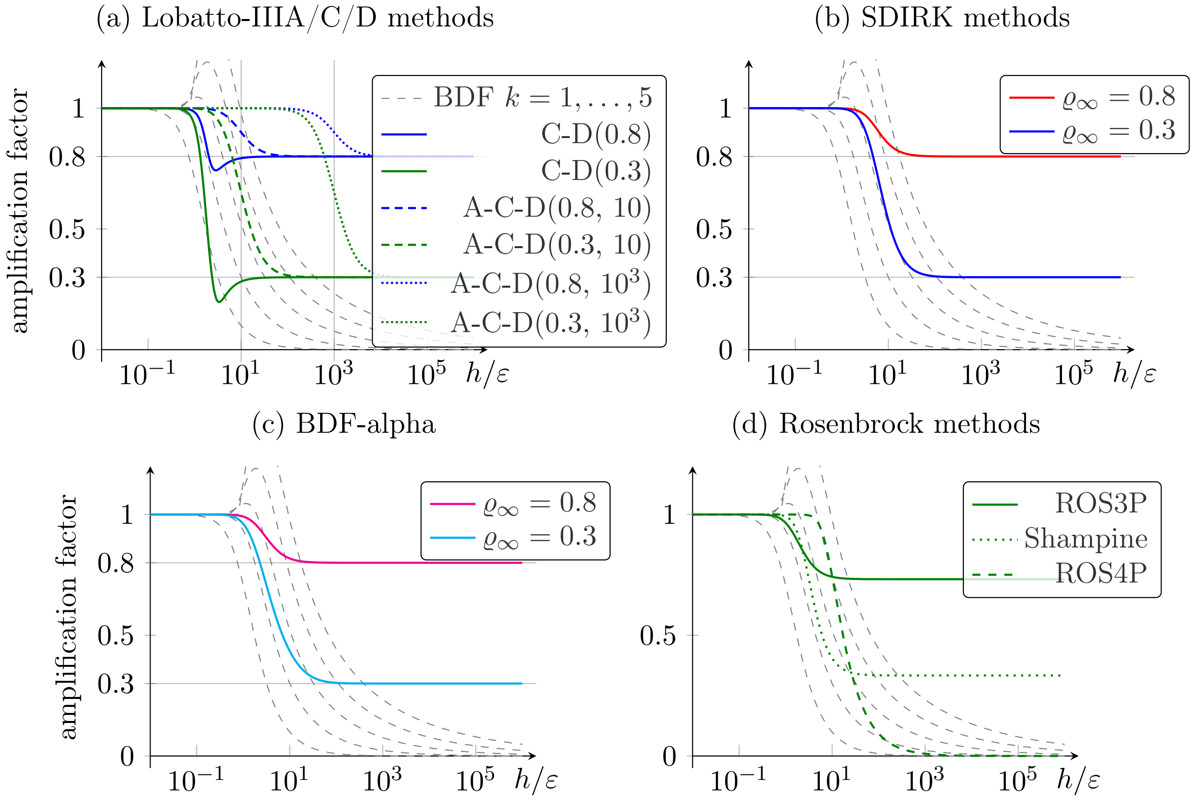
(e) Further methods with improved dissipation behavior

Although typically not explicitly regarded as methods with user-adaptive damping, linearly implicit methods *should be noted among the methods with optimized damping properties. These methods inherit the good stability properties of the underlying Runge–Kutta formulae while their computational effort per time step is usually not only smaller but also almost constant such that they naturally appear to be the method of choice for real-time simulations. The work of Strehmel and Weiner (1989), Hairer et al. (1989b) and Scholz (1989) laid the theoretical foundations for the analysis of these methods in the SPP setting. Recent results and a comprehensive overview for the application in case of stiff mechanical systems can be found in (Simeon, 2013, Becker et al., 2014, Becker, 2012).*

To give a fair sample, we choose the methods of Shampine (1982) with $\varrho_\infty = \frac{1}{3}$, the code ROS3P of Lang and Verwer (2001) ($\varrho_\infty = |1-\sqrt{3}| \approx 0.732$) and Rodasp of Steinebach and Rentrop (2001). For the latter two, the 'p' indicates an optimization with respect to the Prothero–Robinson test example.

Still within the framework of Runge–Kutta methods we also mention the method of Bazzi and Anderheggen (1982) which was one of the first methods in structural dynamics highlighting the importance of numerical damping from one to zero and may be written as a Runge–Kutta–Nyström method.

To compromise between numerical damping and energy consistency, Orden and Romero (2012) proposed so-called 'energy-entropy-momentum integration methods'. These algorithms damp out high frequency oscillations while still preserving symmetries in the model. At last, there exists also a series of on-the-fly filtering-techniques and waveform methods acting directly in the Fourier space of the variables but they are mostly constructed according to specific problems and there is yet no unified theory as for the Newmark integrators in the linear regime.



(a) Lobatto-IIIA/C/D methods

(b) SDIRK methods

(c) BDF-alpha

(d) Rosenbrock methods

Figure 4.9: Numerical damping of some methods from Remark 4.22

4.2.3 Initializing the acceleration-like variables

As overshoot comprises short-term error amplification *and* the initial excitation, and we have emphasized the importance of requirement (d) from Remark 4.19 that methods from structural dynamics should be self starting, the choice of the acceleration-like variable $\boldsymbol{a}_0$ plays an important role.

The original setting was based on the intuitive fact that $\boldsymbol{a}_n$ simply is an acceleration by its physical units:

$$\boldsymbol{a}_0 := \ddot{\boldsymbol{q}}(t_0) = \dot{\boldsymbol{v}}_0\,. \tag{4.35}$$

Whenever we do not explicitly specify it differently in the remaining part of this thesis we will also use (4.35) to define $\boldsymbol{a}_0$. A more detailed analysis, cf. (5.4) below, shows, however, that for second order approximations $\dot{\boldsymbol{v}}_n$, it holds

$$\boldsymbol{a}_n = \ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h) + \mathcal{O}(h^2)\,, \tag{4.36}$$

which yields a second order approximation to the exact accelerations $\ddot{\boldsymbol{q}}(t_n)$ only for $\Delta_\alpha = 0$, i. e., the 'classical' Newmark algorithm[1]. Fortunately, the algorithms damp out these first-order error terms and do not lower the second order of position and velocity variables. Note, however, that this poor initialization causes a genuine problem if the integration has to be re-initialized, for certain variable time-step implementations, or if the algorithms are combined with projection techniques. Also, in case of constrained mechanical systems, the Lagrange multipliers are defined on the level of accelerations and inherit the order reduction from $\boldsymbol{a}$ and $\dot{\boldsymbol{v}}$.

To remedy the poor transient convergence behavior, there have been different proposals in the literature (Negrut et al., 2005, Lunk and Simeon, 2006, Jay and Negrut, 2008) where, based on (4.36), the selection of $\boldsymbol{a}_0$ is improved. If, by the same or a different method, an approximation $\bar{\boldsymbol{a}}_1 \approx \ddot{\boldsymbol{q}}(t_1)$ has been acquired, one can use

$$\boldsymbol{a}_0 := (1 - \Delta_\alpha)\ddot{\boldsymbol{q}}(t_0) + \Delta_\alpha \bar{\boldsymbol{a}}_1\,.$$

A similar approach has been used by Kettmann (2009) where the first integration step was carried out using the trapezoidal rule. Jay and Negrut (2007) argue that the algorithm in the first step then corresponds to a non-damping algorithm which should be excluded. Even more, to some extend, this violates requirement (d) from Remark 4.19. Arnold et al. (2015a) therefore advise to get the approximation just from the initial values and a Taylor series approximation

$$\boldsymbol{a}_0 := \ddot{\boldsymbol{q}}(t_0) + \boldsymbol{\delta}^{\boldsymbol{a}}_{\mathrm{corr}}\,, \quad \boldsymbol{\delta}^{\boldsymbol{a}}_{\mathrm{corr}} \approx h\Delta_\alpha \dddot{\boldsymbol{q}}(t_0)\,,$$

see Section 6.1 below. In the context of index-3 systems and stiff mechanical systems, all these ideas do not suffice to drastically improve the methods which is due to an order reduction that we will explain in the next chapter.

[1] Note that Δ_α may be negative.

Chapter 5

Error analysis

5.1 Basics of error analysis

There are several ways to approach the analysis of the numerical properties of (4.2) for the application to DAEs and SPPs. The investigations of Erlicher et al. (2002) in the nonlinear ODE case and Yen et al. (1998), Lunk and Simeon (2006), and Arnold and Brüls (2007) for the index-1, index-2, and index-3 case, respectively, were based on the convergence theory for multistep methods (Hairer and Wanner, 2002, Sect. VII.3). To gain a deeper understanding of the transient phase, it has proven useful to rely on a onestep representation of the algorithm (Arnold et al., 2015a). So, the propagation of error terms throughout the time integration will be studied using a coupled recursion of vector-valued sequences, consisting of error terms on levels of position and velocity *and* acceleration-like and multiplier variables (Deuflhard et al., 1987). Since its onestep implementation is one of the main features of (4.2), this can be achieved in a relatively straightforward way.

The onestep recursion in all considered problem classes will in the end lead to a similar structure of the error propagation:

Lemma 5.1 (Recursion of vector valued sequences (Arnold et al., 2016, Theorem 4.16))
For vector valued sequences $(\boldsymbol{E}^{\boldsymbol{y}}_n)_{n\geq 0} \subseteq \mathbb{R}^{n_{\boldsymbol{y}}}$, $(\boldsymbol{E}^{\boldsymbol{z}}_n)_{n\geq 0} \subseteq \mathbb{R}^{n_{\boldsymbol{z}}}$ satisfying

$$\|\boldsymbol{E}^{\boldsymbol{y}}_{n+1}\| \leq \|\boldsymbol{E}^{\boldsymbol{y}}_n\| + Lh(\|\boldsymbol{E}^{\boldsymbol{y}}_{n,n+1}\| + \|\boldsymbol{E}^{\boldsymbol{z}}_{n,n+1}\|) + hM\,, \tag{5.1a}$$

$$\|\boldsymbol{E}^{\boldsymbol{z}}_{n+1} - \mathbf{T}\boldsymbol{E}^{\boldsymbol{z}}_n\| \leq L(\|\boldsymbol{E}^{\boldsymbol{y}}_{n,n+1}\| + h\|\boldsymbol{E}^{\boldsymbol{z}}_{n,n+1}\|) + M \tag{5.1b}$$

with non-negative constants L, M and a matrix $\mathbf{T} \in \mathbb{R}^{n_{\boldsymbol{z}}\times n_{\boldsymbol{z}}}$ with spectral radius $\varrho(\mathbf{T}) < 1$, there exist constants $C > 0$, $\tilde{L} > 0$ and $h_0 \geq 0$, independent of n and h, such that whenever $h \in (0, h_0]$, the estimates

$$\|\boldsymbol{E}^{\boldsymbol{y}}_n\| \leq C\,\mathrm{e}^{\tilde{L}nh}\left(\|\boldsymbol{E}^{\boldsymbol{y}}_0\| + h\|\boldsymbol{E}^{\boldsymbol{z}}_0\| + M\right)\,, \tag{5.2a}$$

$$\|\boldsymbol{E}^{\boldsymbol{z}}_n - \mathbf{T}^n\boldsymbol{E}^{\boldsymbol{z}}_0\| \leq C\,\mathrm{e}^{\tilde{L}nh}\left(\|\boldsymbol{E}^{\boldsymbol{y}}_0\| + h\|\boldsymbol{E}^{\boldsymbol{z}}_0\| + M\right) \tag{5.2b}$$

hold.

The notation $\bullet_{n,n+1}$ shall serve as an abbreviation for error terms on both time levels

$$\|\bullet_{n,n+1}\| = \|\bullet_n\| + \|\bullet_{n+1}\|\,,$$

and will occasionally be used to simplify notations. In the convergence analysis below the vectors $\boldsymbol{E}^{\boldsymbol{y}}_n$, $\boldsymbol{E}^{\boldsymbol{z}}_n$ will comprise of condensed error terms on the different levels of the problem while the constant M will collect additional error terms that stem from higher order estimates, local truncation errors and modeling errors of the singularly perturbed problems compared to the equations of the constrained mechanical systems.

Corollary 5.2 (Arnold et al. (2015a, Lemma 7))
Given the assumptions of Lemma 5.1, the triangle inequality may be used to get the estimate

$$\|\boldsymbol{E}_n^{\boldsymbol{z}}\| \le \|\mathbf{T}^n \boldsymbol{E}_0^{\boldsymbol{z}}\| + C\, \mathrm{e}^{\tilde{L}nh} \left(\|\boldsymbol{E}_0^{\boldsymbol{y}}\| + h\|\boldsymbol{E}_0^{\boldsymbol{z}}\| + M \right) .$$

The proof is based on an error recursion in terms of $\|\boldsymbol{E}_n^{\boldsymbol{y}}\|$ and $\|\boldsymbol{E}_n^{\boldsymbol{z}} - \mathbf{T}^n \boldsymbol{E}_0^{\boldsymbol{z}}\|$ and follows in a similar way as 'the classical' error recursion for coupled error propagation in DAE convergence analysis (Deuflhard et al., 1987). Note that for the mapping to define a contraction in the variables $\boldsymbol{E}_n^{\boldsymbol{z}}$, it is necessary to define a norm such that in the corresponding matrix norm $\|\mathbf{T}\|$ is bounded by one, which relies on $\varrho(\mathbf{T}) < 1$.

Definition of global errors In the (index-2 and index-3) DAE case, the global errors after the n-th time step, $n = 0, 1, \ldots$, are self-evidently defined as the difference of numerical and analytic solution.

$\boxed{\text{DAE-case}}$

$$\boldsymbol{e}_n^{\boldsymbol{q}} := \boldsymbol{q}(t_n) - \boldsymbol{q}_n\,, \qquad \boldsymbol{e}_n^{\boldsymbol{v}} := \dot{\boldsymbol{q}}(t_n) - \boldsymbol{v}_n\,, \quad \boldsymbol{e}_n^{\dot{\boldsymbol{v}}} := \ddot{\boldsymbol{q}}(t_n) - \dot{\boldsymbol{v}}_n\,, \tag{5.3a}$$

$$\boldsymbol{e}_n^{\boldsymbol{a}} := \ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h) - \boldsymbol{a}_n \qquad \text{with, see (4.3), } \Delta_\alpha = \alpha_m - \alpha_f\,. \tag{5.3b}$$

We implicitly assume that $\ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h)$ exists or that the solution can be extended to this time instance such that the local truncation errors and global errors to be specified below are always well-defined. The definition of the error in $\boldsymbol{\lambda}(t)$ is postponed to the corresponding sections.

Numerical solutions of SPPs are denoted by having the perturbation parameter

- $()^\delta$ in the strongly damped case and
- $()^\varepsilon$ for the stiff mechanical systems

as a superscript. Apart from this we keep the notation for all error terms: As outlined in Chapter 3, higher derivatives of the SPP solutions $\boldsymbol{q}^\delta(t)$ and $\boldsymbol{q}^\varepsilon(t)$, respectively, cannot be bounded as it is typically done in the error analysis for nonstiff ODEs. In order to obtain upper bounds, we will exclude these strongly attractive or highly oscillating functions and instead of global errors $\boldsymbol{e}_n^{\boldsymbol{q}} = \boldsymbol{q}^{\delta|\varepsilon}(t) - \boldsymbol{q}_n^{\varepsilon|\delta}$ etc. work with the smooth solutions of the according DAE systems, denoted by $\boldsymbol{q}(t)$ and $\boldsymbol{\lambda}(t)$ such that (5.3a) and (5.3b) may be utilized furthermore:

$\boxed{\text{SPP-case}}$

$$\boldsymbol{e}_n^{\boldsymbol{q}} := \boldsymbol{q}(t_n) - \boldsymbol{q}_n^{\delta|\varepsilon}\,, \qquad \boldsymbol{e}_n^{\boldsymbol{v}} := \dot{\boldsymbol{q}}(t_n) - \boldsymbol{v}_n^{\delta|\varepsilon}\,, \qquad \boldsymbol{e}_n^{\dot{\boldsymbol{v}}} := \ddot{\boldsymbol{q}}(t_n) - \dot{\boldsymbol{v}}_n^{\delta|\varepsilon}\,, \tag{5.3c}$$

$$\boldsymbol{e}_n^{\boldsymbol{a}} := \ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h) - \boldsymbol{a}_n^{\delta|\varepsilon}\,. \tag{5.3d}$$

In view of the Rubin–Ungar Theorem in Corollaries 3.9 and 3.18, this suffices when we assume the following:

Assumption 5.3 (Beyond classical convergence theory).
There exists a constant $C_0 > 0$ that may depend on the parameters of the Newmark method (4.2) and the solution $\boldsymbol{q}(t)$ of the slow system (2.12), such that time step size $h > 0$ and penalty parameters δ and ε fulfill the inequalities

$$\delta < C_0 h\,, \quad \varepsilon < C_0 h\,.$$

With regard to the results in the Runge–Kutta case, cf. Remarks 3.11 and 3.20, this appears as a reasonable assumption. Defining errors with respect to the DAE solutions also bears the advantage that the similarities of the DAE and SPP case are highlighted. It is, nevertheless, a rather unfounded supposal to assume that the initial values of the SPPs fulfill the constraints of the DAE systems. Using the projections from Section 2.1.3, we can still construct a unique correspondence of the initial values of the SPPs and solutions of the DAEs: For initial values sufficiently close to the constraint manifolds, we use $\boldsymbol{\pi}$ and $\mathbf{P} =: \mathbf{P}_0$ to map them onto consistent initial values $(\boldsymbol{q}(t_0), \dot{\boldsymbol{q}}(t_0))^\top$ that define the slow motion of the system. For all following time steps, the projection matrix $\mathbf{P}_n := \mathbf{P}(\boldsymbol{q}(t_n))$, $n \geq 1$, is defined by (2.22) using the argument $\boldsymbol{q} = \boldsymbol{q}(t_n)$, such that we can define new error terms

$$\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{x}} := \mathbf{P}_n \boldsymbol{e}_n^{\boldsymbol{x}}\,, \text{ for } \boldsymbol{x} \in \{\boldsymbol{q}, \boldsymbol{v}, \dot{\boldsymbol{v}}, \boldsymbol{a}\}\,. \tag{5.3e}$$

Note that $\mathbf{P}_n$ is defined by $\boldsymbol{q}(t_n)$, which depends *just* on the initial values: For $n \geq 1$, it holds in general $\boldsymbol{\pi}\boldsymbol{q}_n \neq \boldsymbol{q}(t_n)$ in the sense of the definition from (2.21) and (2.22) as $\mathbf{P}_n \boldsymbol{v}_n$ is no longer the mass-orthogonal projection of $\boldsymbol{v}_n$ onto $T_{\boldsymbol{\pi}\boldsymbol{q}_n} \mathfrak{M}^{\mathrm{s}}$. The overall errors $\boldsymbol{e}_n^{\boldsymbol{x}} = \mathbf{P}_n \boldsymbol{e}_n^{\boldsymbol{x}} + (\mathbf{I} - \mathbf{P}_n)\boldsymbol{e}_n^{\boldsymbol{x}}$ comprise also terms in the mass-orthogonal complement, which we will identify by

$$\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{x}} := \mathbf{G}(\boldsymbol{q}(t_n))\boldsymbol{e}_n^{\boldsymbol{x}}\,, \quad \text{for } \boldsymbol{x} \in \{\boldsymbol{q}, \boldsymbol{v}, \dot{\boldsymbol{v}}, \boldsymbol{a}\}\,. \tag{5.3f}$$

Up to multiplication with an invertible matrix, this is in fact the orthogonal component to the manifold because $\mathbf{I} - \mathbf{P}_n = [(\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}) \cdot \mathbf{G}](\boldsymbol{q}(t_n))$.

Local truncation errors Also for the definition of local truncation errors, we will only use smooth solutions, i. e., those of the index-3 DAE system with projected initial values. This allows for a Taylor expansion with bounded higher order derivatives and an equivalent treatment in the DAE and the SPP case. On position level, we obtain

$$\begin{aligned}
\boldsymbol{l}_n^{\boldsymbol{q}} :&= \boldsymbol{q}(t_{n+1}) - \big(\boldsymbol{q}(t_n) + h\dot{\boldsymbol{q}}(t_n) + h^2(\tfrac{1}{2} - \beta)\ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h) + h^2\beta\ddot{\boldsymbol{q}}(t_{n+1} + \Delta_\alpha h)\big) &\quad (5.4a)\\
&= \boldsymbol{q}(t_n) + h\dot{\boldsymbol{q}}(t_n) + \frac{h^2}{2}\ddot{\boldsymbol{q}}(t_n) + \frac{h^3}{6}\dddot{\boldsymbol{q}}^0(t_n) + \mathcal{O}(h^4) \\
&\quad - \boldsymbol{q}(t_n) - h^2(\tfrac{1}{2} - \beta)\left(\ddot{\boldsymbol{q}}(t_n) + \Delta_\alpha h\dddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^2)\right) \\
&\quad - h^2\beta\left(\ddot{\boldsymbol{q}}(t_n) + (h + \Delta_\alpha h)\dddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^2)\right) \\
&= \frac{h^3}{6}\left(1 - 6\beta - 3(\alpha_m - \alpha_f)\right)\dddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^4)\,. &\quad (5.4b)
\end{aligned}$$

To simplify the later representation, we follow Arnold and Brüls (2007) and define the operator

$$\boldsymbol{\Delta}_h(\cdot)_n := \frac{(\cdot)_{n+1} - (\cdot)_n}{h}$$

of forward finite differences and note as a first result (independent of whether the ODE or the DAE case is analyzed) that (5.4b) implies

$$\boldsymbol{\Delta}_h\, \boldsymbol{l}_n^{\boldsymbol{q}} = \frac{\boldsymbol{l}_{n+1}^{\boldsymbol{q}} - \boldsymbol{l}_n^{\boldsymbol{q}}}{h} = \mathcal{O}(h^3)\,.$$

The local truncation error for the velocity coordinates $\boldsymbol{v}_n$ is defined and analyzed in a similar fashion. For it to be of order three, the second order consistency condition (4.3) needs to be fulfilled. It holds

$$\begin{aligned}
\boldsymbol{l}_n^{\boldsymbol{v}} :&= \dot{\boldsymbol{q}}(t_{n+1}) - \left(\dot{\boldsymbol{q}}(t_n) + h(1-\gamma)\ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h) + h\gamma\ddot{\boldsymbol{q}}(t_{n+1} + \Delta_\alpha h)\right) \\
&= \dot{\boldsymbol{q}}(t_n) + h\ddot{\boldsymbol{q}}(t_n) + \frac{h^2}{2}\dddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^3) - \dot{\boldsymbol{q}}(t_n) - h(1-\gamma)\ddot{\boldsymbol{q}}(t_n) \\
&\quad - h^2(1-\gamma)\Delta_\alpha\dddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^3) - h\gamma\ddot{\boldsymbol{q}}(t_n) - h\gamma(h + h\Delta_\alpha)\dddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^3) \\
&= h^2\left(\tfrac{1}{2} - (1-\gamma)\Delta_\alpha - \gamma - \gamma\Delta_\alpha\right)\dddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^3) \\
&= \mathcal{O}(h^3) \text{ if } \gamma = \frac{1}{2} - \Delta_\alpha\,.
\end{aligned}$$

Also for the local truncation errors on acceleration level, we take into account that the acceleration-like variables are approximations to $\ddot{\boldsymbol{q}}$ at shifted time instances.

$$\begin{aligned}
\boldsymbol{l}_n^{\boldsymbol{a}} :&= (1-\alpha_m)\ddot{\boldsymbol{q}}(t_{n+1} + \Delta_\alpha h) + \alpha_m\ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h) - (1-\alpha_f)\ddot{\boldsymbol{q}}(t_{n+1}) - \alpha_f\ddot{\boldsymbol{q}}(t_n) \\
&= (1-\alpha_m)\ddot{\boldsymbol{q}}(t_n) + (1-\alpha_m)(1+\Delta_\alpha)h\dddot{\boldsymbol{q}}(t_n) + \alpha_m\ddot{\boldsymbol{q}}(t_n) + \alpha_m\Delta_\alpha h\dddot{\boldsymbol{q}}(t_n) \\
&\quad - (1-\alpha_f)\ddot{\boldsymbol{q}}(t_n) - (1-\alpha_f)h\dddot{\boldsymbol{q}}(t_n) - \alpha_f\ddot{\boldsymbol{q}}(t_n) + \mathcal{O}(h^2) \\
&= \mathcal{O}(h^2)\,.
\end{aligned}$$

Note that from the given definition of the local truncation errors $\boldsymbol{l}_n^{\boldsymbol{a}}$ it follows

$$(1-\alpha_m)\boldsymbol{e}_{n+1}^{\boldsymbol{a}} + \alpha_m\boldsymbol{e}_n^{\boldsymbol{a}} = (1-\alpha_f)\boldsymbol{e}_{n+1}^{\dot{\boldsymbol{v}}} + \alpha_f\boldsymbol{e}_n^{\dot{\boldsymbol{v}}} + \boldsymbol{l}_n^{\boldsymbol{a}}\,, \tag{5.5}$$

relating the global errors on acceleration level. Note also that the estimate $\boldsymbol{l}_n^{\boldsymbol{a}} = \mathcal{O}(h^2)$ is independent of the choice of parameters α_f, α_m. If we were to construct methods of order three as in Example 4.17 (c), just equating the error constants of the third powers of h would not suffice. Instead, one would have to eliminate the acceleration-like variables first (Erlicher et al. (2002), Kettmann (2009)) or explicitly compute the second order error constants $C_{\boldsymbol{a}}$ in

$$\boldsymbol{e}_n^{\boldsymbol{a}} = \ddot{\boldsymbol{q}}(t_n) + \Delta_\alpha h\dddot{\boldsymbol{q}}(t_n) + C_{\boldsymbol{a}}h^2\ddddot{\boldsymbol{q}}(t_n) - \boldsymbol{a}_n + \mathcal{O}(h^3)\,.$$

Before we turn our attention to the DAE and SPP analysis, we start by several elementary consequences of the definition of global and local errors which do not depend on the specific problem structure.

As a first step, we observe that the difference of the projection matrices $\mathbf{P}_n$ from Section 2 can be estimated as follows using Taylor series expansion

$$\begin{aligned}
\mathbf{P}_{n+1} - \mathbf{P}_n &= \left[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\mathbf{G}\right](\boldsymbol{q}(t_n)) - \left[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\mathbf{G}\right](\boldsymbol{q}(t_{n+1})) \\
&= -h\cdot\frac{\partial(\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\mathbf{G})}{\partial\boldsymbol{q}}(\cdot, \dot{\boldsymbol{q}}(t_n)) + \mathcal{O}(h^2)
\end{aligned}$$

and in particular $\mathbf{P}_{n+1} - \mathbf{P}_n = \mathcal{O}(h)$.

Lemma 5.4
It holds

$$\begin{aligned}
\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{q}} &= \boldsymbol{e}_n^{\boldsymbol{v}} + h(1/2 - \beta)\boldsymbol{e}_n^{\boldsymbol{a}} + h\beta\boldsymbol{e}_{n+1}^{\boldsymbol{a}} + \frac{\boldsymbol{l}_n^{\boldsymbol{q}}}{h}, \\
\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{v}} &= (1-\gamma)\boldsymbol{e}_n^{\boldsymbol{a}} + \gamma\boldsymbol{e}_{n+1}^{\boldsymbol{a}} + \frac{\boldsymbol{l}_n^{\boldsymbol{v}}}{h}\,.
\end{aligned}$$

Proof. The claims follow directly from the definition of the involved error terms. □

Corollary 5.5
The global errors on position level and projected global errors on velocity level obey the following onestep recursions

$$
\begin{aligned}
\boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\boldsymbol{q}} &= \mathcal{O}(1)\left(\|\boldsymbol{e}_n^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{a}}\| + h\|\boldsymbol{e}_{n+1}^{\boldsymbol{a}}\|\right) + \mathcal{O}(h^2)\,, && \text{(5.6a)}\\
\mathbf{G}(\boldsymbol{q}(t_n))\, \boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\boldsymbol{q}} &= \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} + h(\tfrac{1}{2}-\beta)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + h\beta\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \tfrac{1}{h}\mathbf{G}(\boldsymbol{q}(t_n))\boldsymbol{l}_n^{\boldsymbol{q}} + \mathcal{O}(h^2)\|\boldsymbol{e}_{n+1}^{\boldsymbol{a}}\| && \text{(5.6b)}\\
\boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}} &= (1-\gamma)\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} + \gamma\boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} + \mathcal{O}(1)(\|\boldsymbol{e}_n^{\boldsymbol{v}}\| + \tfrac{1}{h}\|\boldsymbol{l}_n^{\boldsymbol{v}}\|) + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{a}}\|\,, && \text{(5.6c)}\\
\boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} &= (1-\gamma)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + \gamma\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \mathcal{O}(1)(\|\boldsymbol{e}_n^{\boldsymbol{v}}\| + \tfrac{1}{h}\|\boldsymbol{l}_n^{\boldsymbol{v}}\|) + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{a}}\|\,. && \text{(5.6d)}
\end{aligned}
$$

Proof. Again, the assertions are a simple deduction (from the estimates of Lemma 5.4) taking into account that the solution $\boldsymbol{q}(t)$ is sufficiently smooth such that $\mathbf{P}_{n+1} - \mathbf{P}_n = \mathcal{O}(h)$ and the same estimate holds for $\mathbf{G} = \mathbf{G}(\boldsymbol{q}(t))$ instead of $\mathbf{P}$. In particular, the last equation follows from

$$
\begin{aligned}
\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{v}} - \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} &= \mathbf{G}(\boldsymbol{q}(t_{n+1}))\cdot(\boldsymbol{e}_{n+1}^{\boldsymbol{v}} - \boldsymbol{e}_n^{\boldsymbol{v}}) + (\mathbf{G}(\boldsymbol{q}(t_{n+1})) - \mathbf{G}(\boldsymbol{q}(t_n)))\,\boldsymbol{e}_n^{\boldsymbol{v}}\\
&= \mathbf{G}(\boldsymbol{q}(t_{n+1}))\cdot(h(1-\gamma)\boldsymbol{e}_n^{\boldsymbol{a}} + h\gamma\boldsymbol{e}_{n+1}^{\boldsymbol{a}} + \boldsymbol{l}_n^{\boldsymbol{v}}) + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{v}}\|\\
&= h(1-\gamma)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + h\gamma\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \mathcal{O}(h^2)\|\boldsymbol{e}_n^{\boldsymbol{a}}\| + \mathcal{O}(1)\|\boldsymbol{l}_n^{\boldsymbol{v}}\| + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{v}}\|\,.
\end{aligned}
$$

□

As has already been outlined in the beginning of this chapter, the error analysis for the Newmark integration family will be carried out using a vector-valued coupled onestep error recursion connecting the global errors $(\boldsymbol{e}_n^*)_{n\geq 0}$ and their projected counterparts with respect to the constraint manifold (Deuflhard et al., 1987). The analysis of the algorithm in the singularly perturbed settings is guided by the corresponding DAE analysis, so we will give the proof for the index-2 and index-3 case first and later put emphasize on the crucial points where SPP and DAE case differ significantly.

5.2 The DAE case

In Chapter 2 we saw that there are several ways to tackle the equations of motion for constrained mechanical systems numerically. Yen et al. (1998) base an integrator with the basic structure of (4.2a), (4.2b), (4.2c) on the index-1 formulation (2.14) and a Gear–Gupta–Leimkuhler-like stabilization technique to include velocity constraints as well. The algorithm is proven to be second order accurate for the position variables and first order on velocity level. The proposals of Lunk and Simeon (2006) and Jay and Negrut (2007, 2008, for HHT and CH(ϱ_∞), respectively) introduced a new algorithmic parameter such that Lagrange multipliers from previous steps enter the integration procedure. For certain parameter values, the method thus becomes equivalent to a Lobatto-IIIA/IIIB pair. Both algorithms use a stabilized version of the index-2 formulation. The latter proposal also deals with nonholonomic constraints and is stated as an overdetermined system. It is shown that both (stabilized) algorithms have second order of convergence in $\boldsymbol{q}$ and $\boldsymbol{v}$ and first order in $\boldsymbol{\lambda}$. Using an adjusted initialization for $\boldsymbol{a}_0$, the order for Lagrange multipliers is also two (Jay and Negrut, 2008).

In Remark 4.3 we presented some of the various ways of extending the ODE algorithms to the case of constrained systems and outlined the advantages of enforcing the constraints at each time step directly, compare (4.2d). For HHT, this concept goes back to the investigations of Cardona and Géradin (1989) and has been extended to the CH(ϱ_∞) case by Brüls (2005) and the stabilized index-2 form by Arnold (2009).

The equilibrium equations of the Newmark algorithm (4.2d) take the form

$$\mathbf{M}(\boldsymbol{q}_n)\dot{\boldsymbol{v}}_n = \boldsymbol{f}(\boldsymbol{q}_n, \boldsymbol{v}_n) - \mathbf{G}^\top(\boldsymbol{q}_n)\boldsymbol{\lambda}_n\,, \tag{5.7a}$$

$$\mathbf{0} = \begin{cases} \mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n & \text{for the index-2 case and} \\ \boldsymbol{g}(\boldsymbol{q}_n) & \text{for the index-3 case.} \end{cases} \tag{5.7b}$$

In the DAE setting, the definition of global errors in $\boldsymbol{\lambda}$ is straightforward

$$\boldsymbol{e}_n^{\boldsymbol{\lambda}} := \boldsymbol{\lambda}(t_n) - \boldsymbol{\lambda}_n\,. \tag{5.8}$$

As for the projected error terms and the ones orthogonal to the constraint manifold in (5.3e) and (5.3f), we also introduce the additional error terms

$$\boldsymbol{e}_n^{\mathbf{X}\boldsymbol{\lambda}} := \mathbf{X} \cdot \boldsymbol{e}_n^{\boldsymbol{\lambda}} \tag{5.9}$$

for matrix-valued factors $\mathbf{X} \in \mathbb{R}^{k \times n_{\boldsymbol{\lambda}}}$, $k \in \mathbb{N}$.

As is common in the error analysis of DAE systems, we start the investigations with the imposition of weak estimates on the global errors of all involved coordinates. These estimates will be used to obtain recursion formulae of error terms which finally lead to *stronger* error bounds. The latter ones provide a later justification of the original assumption which can be shown by induction, see (Hairer and Wanner, 2002, part (c) of the proof of Theorem VII.3.5).

Assumption 5.6 (Technical assumption, the DAE case)**.**
For the error analysis in the DAE case we will suppose that there exist constants $C, h_0 > 0$, such that whenever $0 < h \leq h_0$ holds, we have the estimates

$$\|\boldsymbol{e}_m^{\boldsymbol{q}}\| \leq Ch\,, \quad \|\boldsymbol{e}_m^{\boldsymbol{v}}\| + \|\boldsymbol{e}_m^{\boldsymbol{a}}\| + \|\boldsymbol{e}_m^{\boldsymbol{\lambda}}\| \leq C \tag{5.10}$$

for all $m \geq 0$, $t_0 + mh \leq t_{\text{end}}$.

Lemma 5.7 (Errors on acceleration level (Arnold et al., 2015a, Lemma 3))
For the errors $\boldsymbol{e}_n^{\dot{\boldsymbol{v}}}$, $\boldsymbol{e}_n^{\boldsymbol{\lambda}}$ on acceleration level the following estimates hold

$$\boldsymbol{e}_n^{\dot{\boldsymbol{v}}} = \mathcal{O}(1)(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\|) - \boldsymbol{e}_n^{\mathbf{M}^{-1}\mathbf{G}^\top\boldsymbol{\lambda}}\,,$$

where the argument of $\mathbf{M}^{-1}\mathbf{G}^\top$ in the definition of $\boldsymbol{e}_n^{\mathbf{M}^{-1}\mathbf{G}^\top\boldsymbol{\lambda}}$ is to be taken at $\boldsymbol{q}(t_n)$.

Proof. We multiply the equilibrium condition of the dynamic equations (2.12) by $\mathbf{M}^{-1}(\boldsymbol{q}(t_n))$ and the corresponding numerical equilibrium condition (5.7a) by $\mathbf{M}^{-1}(\boldsymbol{q}_n)$, respectively. Subtraction results in

$$\begin{aligned} \boldsymbol{e}_n^{\dot{\boldsymbol{v}}} + \boldsymbol{e}_n^{\mathbf{M}^{-1}\mathbf{G}^\top\boldsymbol{\lambda}} &= \ddot{\boldsymbol{q}}(t_n) - \dot{\boldsymbol{v}}_n + [\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}(t_n))(\boldsymbol{\lambda}(t_n) - \boldsymbol{\lambda}_n) \\ &= [\mathbf{M}^{-1}\boldsymbol{f}](\boldsymbol{q}(t_n), \dot{\boldsymbol{q}}(t_n)) - [\mathbf{M}^{-1}\boldsymbol{f}](\boldsymbol{q}_n, \boldsymbol{v}_n) \\ &\qquad + ([\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}_n) - [\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}(t_n)))\boldsymbol{\lambda}_n\,, \end{aligned}$$

which already gives the claim, provided that the technical assumption on $\boldsymbol{e}_n^{\boldsymbol{\lambda}}$ from (5.10) holds. □

Using the result of Lemma 5.7, we obtain the following estimates for their corresponding projections onto the tangential and $\mathbf{M}$-orthogonal direction of $\mathfrak{M}^{\mathrm{s}} = \{\boldsymbol{q} : \boldsymbol{g}(\boldsymbol{q}) = \mathbf{0}\}$:

$$\boldsymbol{e}_n^{\mathbf{P}\dot{\boldsymbol{v}}} = \mathcal{O}(1)\left(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\|\right)\,, \tag{5.11a}$$

$$\boldsymbol{e}_n^{\mathbf{G}\dot{\boldsymbol{v}}} = \mathcal{O}(1)\left(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\|\right) - \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}\,. \tag{5.11b}$$

These results simply stem from the definition of $\mathbf{P}_n$ and $\mathbf{S}(\boldsymbol{q}(t_n))$, cf. (2.24) and (2.20). Utilizing the acceleration updates of the algorithm, this leads to error recursions for the errors in acceleration-like and algebraic variables. More precisely, we get:

Lemma 5.8 (Error coupling in $\boldsymbol{a}$ and $\boldsymbol{\lambda}$ (Arnold et al., 2015a, Lemma 5))
The projections of global errors meet the recursions

$$\begin{aligned}(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} &= \mathcal{O}(1)\left(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{e}_{n,n+1}^{\boldsymbol{v}}\|\right) + \mathcal{O}(h)\left(\|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\| + \|\boldsymbol{e}_n^{\boldsymbol{a}}\|\right) + \mathcal{O}(h^2)\,,\\ (1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} &+ (1-\alpha_f)\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda}} + \alpha_f \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}\\ &= \mathcal{O}(1)\left(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{e}_{n,n+1}^{\boldsymbol{v}}\|\right) + \mathcal{O}(h)\left(\|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\| + \|\boldsymbol{e}_n^{\boldsymbol{a}}\|\right) + \mathcal{O}(h^2)\,.\end{aligned}$$

Proof. We prove only the first estimate as the second follows in exactly the same way considering the different values in (5.11). The assertion is a direct consequence of Lemma 5.7 and (5.5).

$$\begin{aligned}(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} &= \mathbf{P}_{n+1}((1-\alpha_m)\boldsymbol{e}_{n+1}^{\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\boldsymbol{a}}) + (\mathbf{P}_n - \mathbf{P}_{n+1})\alpha_m \boldsymbol{e}_n^{\boldsymbol{a}}\\ &= \mathbf{P}_{n+1}((1-\alpha_f)\boldsymbol{e}_{n+1}^{\dot{\boldsymbol{v}}} + \alpha_f \boldsymbol{e}_n^{\dot{\boldsymbol{v}}} + \boldsymbol{l}_n^{\boldsymbol{a}}) + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{a}}\|\\ &= (1-\alpha_f)\boldsymbol{e}_{n+1}^{\mathbf{P}\dot{\boldsymbol{v}}} + \alpha_f \boldsymbol{e}_n^{\mathbf{P}\dot{\boldsymbol{v}}} + \alpha_f(\mathbf{P}_{n+1} - \mathbf{P}_n)\boldsymbol{e}_n^{\dot{\boldsymbol{v}}} + \mathbf{P}_{n+1}\boldsymbol{l}_n^{\boldsymbol{a}} + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{a}}\|\\ &= \mathcal{O}(1)\left(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{e}_{n,n+1}^{\boldsymbol{v}}\|\right) + \mathcal{O}(h)\left(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| + \|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\| + \|\boldsymbol{e}_n^{\boldsymbol{a}}\|\right) + \mathbf{P}_{n+1}\boldsymbol{l}_n^{\boldsymbol{a}}\,,\end{aligned}$$

such that the claim follows from (5.11) and $\mathbf{P}_{n+1}\boldsymbol{l}_n^{\boldsymbol{a}} = \mathcal{O}(h^2)$. Note that the latter one is independent of the parameter choice of the method and that Assumption 5.6 was necessary to derive this result. □

After studying the equilibrium conditions of the force balance in (2.12), we now point our attention to the constraint residuals. Independent of the integrator (index-2 or index-3), we have the following result.

Lemma 5.9 (Constraint residuals and position errors in normal direction (Arnold et al., 2015a, Lemma 4))
The error components in normal direction $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{q}}$ and the residuals in the position constraint relate like

$$\boldsymbol{g}(\boldsymbol{q}_n) = -\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{q}} + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{q}}\|\,.$$

Their difference for two consecutive time steps fulfills

$$-\boldsymbol{\Delta}_h\,\boldsymbol{g}(\boldsymbol{q}_n) = \mathbf{G}(\boldsymbol{q}(t_n))\,\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{q}} + \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_n^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_n)) + \mathcal{O}(h)(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{q}}\|)\,.$$

Proof. We have

$$\begin{aligned}\boldsymbol{g}(\boldsymbol{q}_n) = \boldsymbol{g}(\boldsymbol{q}(t_n) - 1\cdot\boldsymbol{e}_n^{\boldsymbol{q}}) - \overbrace{\boldsymbol{g}(\boldsymbol{q}(t_n) - 0\cdot\boldsymbol{e}_n^{\boldsymbol{q}})}^{=\mathbf{0}} &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\vartheta}\boldsymbol{g}(\boldsymbol{q}(t_n) - \vartheta\boldsymbol{e}_n^{\boldsymbol{q}})\,\mathrm{d}\vartheta\\ = \int_0^1 -\mathbf{G}(\boldsymbol{q}(t_n) - \vartheta\boldsymbol{e}_n^{\boldsymbol{q}})\boldsymbol{e}_n^{\boldsymbol{q}}\,\mathrm{d}\vartheta &= \int_0^1 -\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{q}} + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{q}}\|\,\mathrm{d}\vartheta\,,\end{aligned}$$

and so the first assertion. For the finite difference term, we proceed as

$$\begin{aligned}
-\,\boldsymbol{\Delta}_h\, \boldsymbol{g}(\boldsymbol{q}_n) &= -\frac{\boldsymbol{g}(\boldsymbol{q}_{n+1}) - \boldsymbol{g}(\boldsymbol{q}(t_{n+1})) - (\boldsymbol{g}(\boldsymbol{q}_n) - \boldsymbol{g}(\boldsymbol{q}(t_n)))}{h} \\
&= -\frac{1}{h}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}\vartheta}\boldsymbol{g}(\boldsymbol{q}(t_{n+1}) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_{n+1})\ \mathrm{d}\vartheta + \frac{1}{h}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}\vartheta}\boldsymbol{g}(\boldsymbol{q}(t_n) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_n)\ \mathrm{d}\vartheta \\
&= \int_0^1 \mathbf{G}(\boldsymbol{q}(t_{n+1}) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_{n+1})\, \boldsymbol{\Delta}_h\, \boldsymbol{e}^{\boldsymbol{q}}_n\ \mathrm{d}\vartheta \\
&\qquad + \frac{1}{h}\int_0^1 \big(\mathbf{G}(\boldsymbol{q}(t_{n+1}) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_{n+1}) - \mathbf{G}(\boldsymbol{q}(t_n) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_n)\big)\, \boldsymbol{e}^{\boldsymbol{q}}_n\ \mathrm{d}\vartheta\,.
\end{aligned}$$

For the second summand, we use the Fundamental Theorem of Calculus again: Define

$$\boldsymbol{e}^{\vartheta,\boldsymbol{q}}_n := \boldsymbol{q}(t_{n+1}) - \boldsymbol{q}(t_n) - \vartheta(\boldsymbol{e}^{\boldsymbol{q}}_{n+1} - \boldsymbol{e}^{\boldsymbol{q}}_n) = h\dot{\boldsymbol{q}}(t_n) - \vartheta h\, \boldsymbol{\Delta}_h\, \boldsymbol{e}^{\boldsymbol{q}}_n + \mathcal{O}(h^2)\,,$$

such that for the second term, we get

$$\begin{aligned}
&\frac{1}{h}\int_0^1 \Big(\mathbf{G}(\boldsymbol{q}(t_n) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_n + 1\cdot \boldsymbol{e}^{\vartheta,\boldsymbol{q}}_n) - \mathbf{G}(\boldsymbol{q}(t_n) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_n + 0\cdot \boldsymbol{e}^{\vartheta,\boldsymbol{q}}_n)\Big)\, \boldsymbol{e}^{\boldsymbol{q}}_n\ \mathrm{d}\vartheta \\
&\quad = \frac{1}{h}\int_0^1\int_0^1 \mathsf{R}(\boldsymbol{q}(t_n) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_n + \bar{\vartheta}\boldsymbol{e}^{\vartheta,\boldsymbol{q}}_n)(\boldsymbol{e}^{\boldsymbol{q}}_n, \boldsymbol{e}^{\vartheta,\boldsymbol{q}}_n)\ \mathrm{d}\bar{\vartheta}\ \mathrm{d}\vartheta \\
&\quad = \int_0^1\int_0^1 \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}^{\boldsymbol{q}}_n, \dot{\boldsymbol{q}}(t_n))\ \mathrm{d}\bar{\vartheta}\ \mathrm{d}\vartheta + \mathcal{O}(1)\left(h\|\boldsymbol{\Delta}_h\, \boldsymbol{e}^{\boldsymbol{q}}_n\| + \|\boldsymbol{e}^{\boldsymbol{q}}_n\|\cdot \max_{\vartheta}\|\boldsymbol{e}^{\vartheta,\boldsymbol{q}}_n\|\right) \\
&\quad = \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}^{\boldsymbol{q}}_n, \dot{\boldsymbol{q}}(t_n)) + \mathcal{O}(h)(\|\boldsymbol{e}^{\boldsymbol{q}}_n\| + \|\boldsymbol{\Delta}_h\, \boldsymbol{e}^{\boldsymbol{q}}_n\|)\,,
\end{aligned}$$

where the technical assumption (5.10) on the position errors has been used. The same technique may be used to estimate the first summand. □

Lemma 5.10 (A bound for $\boldsymbol{\Delta}_h\, \boldsymbol{e}^{\mathbf{G}\boldsymbol{v}}_n$)
The finite difference of normal global velocity error components are bounded like

$$\boldsymbol{\Delta}_h\, \boldsymbol{e}^{\mathbf{G}\boldsymbol{v}}_n = \frac{1}{h}\big(\mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n - \mathbf{G}(\boldsymbol{q}_{n+1})\boldsymbol{v}_{n+1}\big) + \mathcal{O}(1)\left(\|\boldsymbol{e}^{\boldsymbol{v}}_n\| + \|\boldsymbol{e}^{\boldsymbol{v}}_{n+1}\| + \|\boldsymbol{\Delta}_h\, \boldsymbol{e}^{\boldsymbol{q}}_n\|\right)\,.$$

Proof. As in the proof of Lemma 5.9, we use an integral representation to estimate $\boldsymbol{e}^{\mathbf{G}\boldsymbol{v}}_n$:

$$\begin{aligned}
\boldsymbol{e}^{\mathbf{G}\boldsymbol{v}}_n &= -\mathbf{G}(\boldsymbol{q}(t_n))\boldsymbol{v}_n = -\mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n + (\mathbf{G}(\boldsymbol{q}(t_n) - 1\cdot \boldsymbol{e}^{\boldsymbol{q}}_n) - \mathbf{G}(\boldsymbol{q}(t_n) - 0\cdot \boldsymbol{e}^{\boldsymbol{q}}_n))\,\boldsymbol{v}_n \\
&= -\mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n - \int_0^1 \mathsf{R}(\boldsymbol{q}(t_n) - \vartheta \boldsymbol{e}^{\boldsymbol{q}}_n)(\boldsymbol{v}_n, \boldsymbol{e}^{\boldsymbol{q}}_n)\ \mathrm{d}\vartheta\,,
\end{aligned}$$

and so the assertion. The $\|\boldsymbol{e}^{\boldsymbol{v}}_{\cdot}\|$-terms enters if the first argument in the bilinear form R is exchanged for the analytic solution $\dot{\boldsymbol{q}}(t)$. □

5.2.1 The index-2 case

For an easier reference we state once more the algorithm in the index-2 setting.

$$\begin{aligned}
\boldsymbol{q}_{n+1} &= \boldsymbol{q}_n + h\boldsymbol{v}_n + h^2(\tfrac{1}{2} - \beta)\boldsymbol{a}_n + h^2\beta\boldsymbol{a}_{n+1}\,, \\
\boldsymbol{v}_{n+1} &= \boldsymbol{v}_n + h(1-\gamma)\boldsymbol{a}_n + h\gamma\boldsymbol{a}_{n+1}\,, \\
(1-\alpha_m)\boldsymbol{a}_{n+1} + \alpha_m\boldsymbol{a}_n &= (1-\alpha_f)\dot{\boldsymbol{v}}_{n+1} + \alpha_f\dot{\boldsymbol{v}}_n\,, \\
\mathbf{M}(\boldsymbol{q}_{n+1})\dot{\boldsymbol{v}}_{n+1} &= \boldsymbol{f}(\boldsymbol{q}_{n+1}, \boldsymbol{v}_{n+1}) - \mathbf{G}^\top(\boldsymbol{q}_{n+1})\boldsymbol{\lambda}_{n+1} \\
\mathbf{G}(\boldsymbol{q}_{n+1})\boldsymbol{v}_{n+1} &= \mathbf{0}
\end{aligned}$$

with $\boldsymbol{q}_0 = \boldsymbol{q}(t_0)$, $\boldsymbol{v}_0 = \dot{\boldsymbol{q}}(t_0)$, $\dot{\boldsymbol{v}}_0 = \ddot{\boldsymbol{q}}(t_0)$ and $\boldsymbol{a}_0 \in \mathbb{R}^{n_q}$ which has to be defined in an initialization phase. Whenever not explicitly stated otherwise we use the choice from (4.35).

Example 5.11 (Error behavior in the index-2 case)
Before we proceed with a detailed analysis of the Newmark integrator in the index-2 case, we consider once again the planar pendulum example from the first chapter.

(a) We used the CH(ϱ_∞)-algorithm with time step size $h = 2 \cdot 10^{-2}$ and obtained in a short transient phase (first ten time steps) the results depicted in Figure 5.1 for the global errors $\boldsymbol{e}_n^{\boldsymbol{\lambda}}$ in the Lagrange multipliers. The main feature of the Chung–Hulbert parameter choice (4.4) lies in controllable numerical damping of the algorithm, so we used $\varrho_\infty = 0$ and—as a comparison—the values $\varrho_\infty = 0.1716, 0.2679, 0.6666$ which should provide less numerical damping.

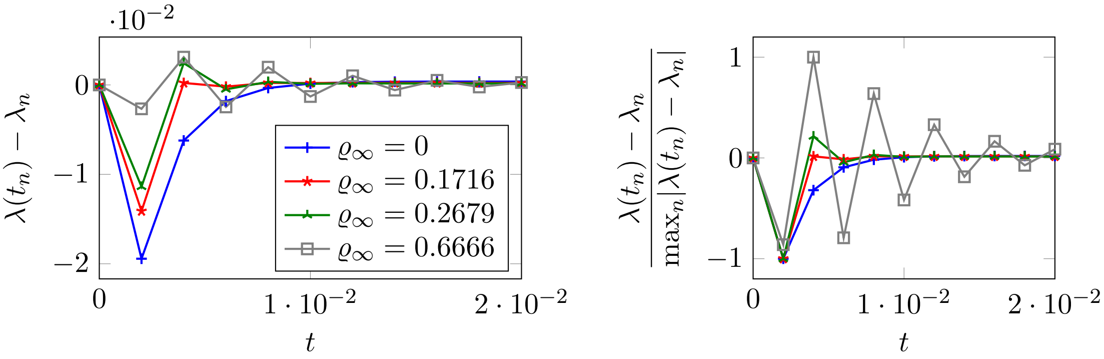


Figure 5.1: Transient error behavior in Lagrange multiplier for the CH(ϱ_∞) method in index-2 formulation, right plot: error scaled with respect to maximal deviation

As the overshoot phenomenon has already been discussed, there is no wonder about the large amplitude in the first few steps. Rather surprising is that it seems that for the larger values of the damping parameter $\varrho_\infty \in \{0.1716, 0.2679\}$, the numerical damping observed in the experiment is in fact stronger(!) Moreover, for $\varrho_\infty = 0$ there are no spurious *oscillations* but instead just a damped term, while for the other settings the sign of the error changes in each of the first integration steps. Both results will become clear once we state the error propagation for the acceleration components in the index-2 case, including a reasoning for the specific choices of ϱ_∞ in this example.

(b) In Figure 5.2 the numerically obtained order of convergence for the pendulum example can be observed. We applied CH(0.75) to the system with two different initial configurations. In the first setting, we have the same situation as in the introductory example from Chapter 1 and find an order reduction. More precisely, the maximum error in acceleration variables is (only) linearly dependent on the time step size h. If we exclude the transient phase and only consider the error for $\{n : t_n \geq 0.4\}$, the order reduction seems to disappear and the expected 'classical' second order is observed. The same holds if we change the initial configuration of the system and start with zero initial velocity $\dot{\boldsymbol{q}}(0) = (0, 0)^\top$: Second order is preserved for the global errors after the transient phase as well as for the first integration steps.

$\diamondsuit$

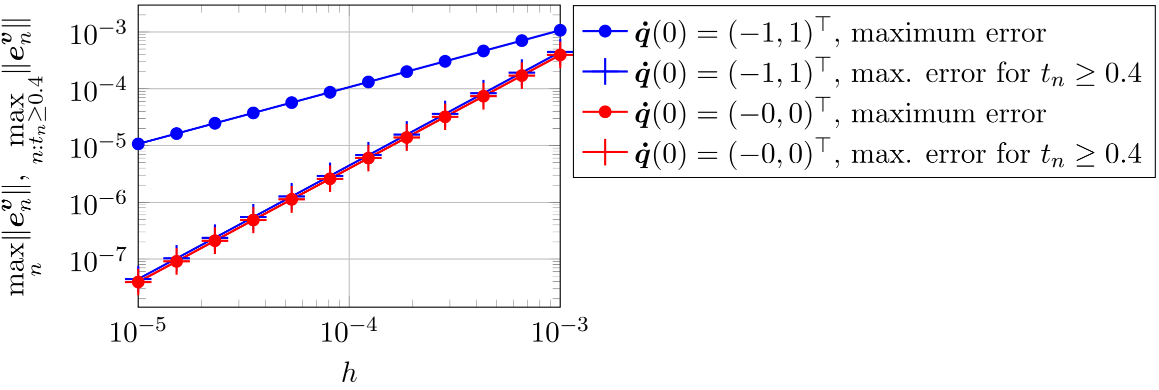


Figure 5.2: Global error of CH(0.75)-method for different initial conditions and with transient phase in-/excluded

Before the two numerical phenomena of Example 5.11 will be explained in the detailed error analysis below, we have a quick look at the errors terms for the acceleration-like variables $\boldsymbol{a}_0$ after initialization.

Lemma 5.12 (Initial error terms (I))
If the initialization of acceleration-like variables is done as originally proposed by Chung and Hulbert (1993), i. e., $\boldsymbol{a}_0 := \dot{\boldsymbol{v}}_0 = \ddot{\boldsymbol{q}}(t_0)$, the initial error terms on acceleration level satisfy

$$\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}} = \mathcal{O}(h)\,, \quad \boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(h)\,.$$

Proof. Taylor expansion gives the result directly. □

The initial global error terms on position and velocity level vanish for the obvious choice $\boldsymbol{q}_0 := \boldsymbol{q}(t_0)$, $\boldsymbol{v}_0 := \dot{\boldsymbol{q}}(t_0)$ as does the error for Lagrange multipliers if the very same ones are initialized by solving (2.15) for the initial values.

We are now prepared to collect all the above estimates for the convergence result in the case of (5.7) with velocity constraints $\mathbf{G}(\boldsymbol{q})\dot{\boldsymbol{q}} = \mathbf{0}$.

Theorem 5.13 (Convergence in the index-2 case)

(a) Let the order condition (4.3) be fulfilled and suppose that the corrector equations are solved such that $\max_n \|\mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n\| = \mathcal{O}(h^3)$. If the starting values satisfy

$$\|\boldsymbol{e}_0^{\boldsymbol{q}}\| + \|\boldsymbol{e}_0^{\boldsymbol{v}}\| = \mathcal{O}(h^2)\,, \quad \|\boldsymbol{e}_0^{\boldsymbol{a}}\| + \|\boldsymbol{e}_0^{\boldsymbol{\lambda}}\| = \mathcal{O}(h)$$

then for the errors of the Newmark integrator for index-2 systems it holds

$$\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| \leq C\,\mathrm{e}^{\tilde{L}(t_n - t_0)}\,h^2\,,$$
$$\left\| \begin{pmatrix} \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} \\ \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}} \\ \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} \end{pmatrix} - \mathbf{T}^n \begin{pmatrix} \boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}} \\ \boldsymbol{e}_0^{\mathbf{S}\boldsymbol{\lambda}} \\ \boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} \end{pmatrix} \right\| \leq C\,\mathrm{e}^{\tilde{L}(t_n - t_0)}\,h^2$$

where $\mathbf{T}$ will be defined in the proof below.

(b) Assume furthermore that

$$\alpha_m < \alpha_f < \tfrac{1}{2} \tag{5.13}$$

holds. If the starting values additionally satisfy

$$\|\boldsymbol{e}_0^{\dot{\boldsymbol{v}}}\| + \|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}\| + \|\mathbf{M}(\boldsymbol{q}_0)\dot{\boldsymbol{v}}_0 - \boldsymbol{f}(\boldsymbol{q}_0,\boldsymbol{v}_0) + \mathbf{G}^\top(\boldsymbol{q}_0)\boldsymbol{\lambda}_0\| = \mathcal{O}(h^{1+\varkappa}) \tag{5.14}$$

for a given constant $\varkappa \in [0,1]$, then the errors on level of Lagrange multipliers are bounded like

$$\|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\| \leq C\left(\varrho^n h^{1+\varkappa} + \mathrm{e}^{\tilde{L}(t_n-t_0)}\, h^2\right),$$

where $0 < \varrho < 1$ is a constant depending on the parameters of the algorithm.

The constants C, $\tilde{L} \geq 0$ are independent of n.

Proof. In preparation of applying Lemma 5.1, we introduce the condensed error vectors $\boldsymbol{E}_n^{\boldsymbol{y}} := (\boldsymbol{e}_n^{\boldsymbol{q}}, \boldsymbol{e}_n^{\boldsymbol{v}})^\top$ and $\boldsymbol{E}_n^{\boldsymbol{z}} := (\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}, \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}, \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}})^\top$ to separate force-related error terms from the ones on position and velocity level. We collect the estimates from Lemma 5.4 to get

$$\boldsymbol{E}_{n+1}^{\boldsymbol{y}} = \begin{pmatrix}\boldsymbol{e}_{n+1}^{\boldsymbol{q}}\\ \boldsymbol{e}_{n+1}^{\boldsymbol{v}}\end{pmatrix} = \begin{pmatrix}\boldsymbol{e}_{n}^{\boldsymbol{q}}\\ \boldsymbol{e}_{n}^{\boldsymbol{v}}\end{pmatrix} + \mathcal{O}(h)\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{a}}\| + \mathcal{O}(1)(\|\boldsymbol{l}_n^{\boldsymbol{q}}\| + \|\boldsymbol{l}_n^{\boldsymbol{v}}\|) = \boldsymbol{E}_n^{\boldsymbol{y}} + \mathcal{O}(h)\|\boldsymbol{E}_{n,n+1}^{\boldsymbol{z}}\| + \mathcal{O}(h^3),$$

where the order condition (4.3) ensures that $\|\boldsymbol{l}_n^{\boldsymbol{v}}\| = \mathcal{O}(h^2)$. In terms of the condensed errors, Lemma 5.8 furthermore provides the two equations

$$\begin{aligned}
(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} &= \mathcal{O}(1)\|\boldsymbol{E}_{n,n+1}^{\boldsymbol{y}}\| + \mathcal{O}(h)\|\boldsymbol{E}_{n,n+1}^{\boldsymbol{z}}\| + \mathcal{O}(h^2),\\
(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} &= -(1-\alpha_f)\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda}} - \alpha_f \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}} + \mathcal{O}(1)\|\boldsymbol{E}_{n,n+1}^{\boldsymbol{y}}\| + \mathcal{O}(h)\|\boldsymbol{E}_{n,n+1}^{\boldsymbol{z}}\| + \mathcal{O}(h^2).
\end{aligned}$$

Note, that $\|\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}\| = \mathcal{O}(1)\|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\|$ and $\|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\| = \mathcal{O}(1)\|\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}\|$ since $\mathbf{S}(\boldsymbol{q})$ is bounded and non-singular. Eventually, we combine Corollary 5.5 and Lemma 5.10 to get

$$\begin{aligned}
(1-\gamma)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + \gamma \boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} &+ \mathcal{O}(1)(\|\boldsymbol{l}_n^{\boldsymbol{v}}\|/h + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{a}}\|)\\
&= \boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}\\
&= \frac{1}{h}(-\mathbf{G}(\boldsymbol{q}_{n+1})\boldsymbol{v}_{n+1} + \mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n) + \mathcal{O}(1)(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{v}}\| + \|\boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\boldsymbol{q}}\|)\\
&= \frac{1}{h}(-\mathbf{G}(\boldsymbol{q}_{n+1})\boldsymbol{v}_{n+1} + \mathbf{G}(\boldsymbol{q}_n)\boldsymbol{v}_n) + \mathcal{O}(1)(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{a}}\| + \frac{1}{h}\|\boldsymbol{l}_n^{\boldsymbol{q}}\|)\\
&= \mathcal{O}(1)\|\boldsymbol{E}_{n,n+1}^{\boldsymbol{y}}\| + \mathcal{O}(h)\|\boldsymbol{E}_{n,n+1}^{\boldsymbol{z}}\| + \mathcal{O}(h^2).
\end{aligned} \tag{5.15}$$

Putting all this together, we arrive at the coupled recursion of Lemma 5.1 and Corollary 5.2 with

$$\mathbf{T} := \begin{pmatrix}(1-\alpha_m)\mathbf{I}_{n_q} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & -\gamma\mathbf{I}_{n_{\boldsymbol{\lambda}}}\\ \mathbf{0} & (1-\alpha_f)\mathbf{I}_{n_{\boldsymbol{\lambda}}} & (1-\alpha_m)\mathbf{I}_{n_{\boldsymbol{\lambda}}}\end{pmatrix}^{-1}\begin{pmatrix}-\alpha_m\mathbf{I}_{n_q} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & (1-\gamma)\mathbf{I}_{n_{\boldsymbol{\lambda}}}\\ \mathbf{0} & -\alpha_f\mathbf{I}_{n_{\boldsymbol{\lambda}}} & -\alpha_m\mathbf{I}_{n_{\boldsymbol{\lambda}}}\end{pmatrix}$$

and $M = \mathcal{O}(h^2)$ and therefore obtain the estimates from part (a) of Theorem 5.13. To obtain (b) we use (5.13) which implies that the error recursion is contractive, i. e., the spectral radius $\varrho(\mathbf{T}) < 1$, see Remark 5.16 below. The explicit structure of $\mathbf{T}$ allows to separate the error terms $\boldsymbol{e}_{n,n+1}^{\mathbf{P}\boldsymbol{a}}$ and $(\boldsymbol{e}_{n,n+1}^{\mathbf{S}\boldsymbol{\lambda}}, \boldsymbol{e}_{n,n+1}^{\mathbf{G}\boldsymbol{a}})^\top$ such that $\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}} = \mathcal{O}(h)$ may be kept while still getting higher order in $\boldsymbol{e}_n^{\boldsymbol{\lambda}}$ as long as (5.14) holds. The constant ϱ may be chosen as any real number with $\varrho(\mathbf{T}_{2:3,2:3}) < \varrho < 1$.

Note that the contractivity condition $\alpha_m < \frac{1}{2}$ implies that the algorithm is zero stable for ODEs, cf. Lemma 4.10. □

Corollary 5.14 (Second order convergence for the index-2 integrator)
Given consistent initial values $(\boldsymbol{q}_0, \boldsymbol{v}_0) = (\boldsymbol{q}(t_0), \dot{\boldsymbol{q}}(t_0))$ and starting with acceleration-like variables satisfying

$$\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}} = \mathcal{O}(h)\,, \quad \boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(h^2)\,, \quad \boldsymbol{e}_0^{\boldsymbol{\lambda}} = \mathcal{O}(h^2)\,,$$

the second order (in the classical ODE setting) Newmark integrators sustain their second order of convergence for the index-2 case, i. e., the global errors on position, velocity and level of Lagrange multipliers are second order provided that (5.13) remains valid. For the original initialization choice $\boldsymbol{a}_0 := \dot{\boldsymbol{v}}_0 = \ddot{\boldsymbol{q}}(t_0)$ the error for the Lagrange multipliers may drop to one. After a transient phase, however, the influence of the first order error terms become negligible.

Proof. The result is a direct consequence of Theorem 5.13. Note that it would not suffice to just use Corollary 5.2 since then the separation of $\|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}\|$ and $\|\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}\|$ up to higher order terms would not be possible. The results of Lemma 5.12 give the second assertion. □

Remark 5.15
We tacitly assumed that the sufficiently exact solution of the corrector equations includes that the update equations for position, velocity and acceleration-like variables in (4.2) *are also realized sufficiently accurate. The implementation proposed in Section 6.1 below will be constructed in such a way that the latter three are fulfilled up to machine precision anyway. Even if the solution of the update equations is carried out in a different way, it is a reasonable assumption that the linear update formulae are at least as well resolved as the nonlinear equations for the equilibrium conditions and the constraints.*

In the original work Arnold et al. (2015a) take the 'tolerance' of the Newton–Raphson algorithm into account: They show that the error estimate of Theorem 5.13 (as well as Theorem 5.21 below) remains valid if the deviation from the constraint manifold defined by $\mathbf{G}(\boldsymbol{q})\dot{\boldsymbol{q}} = \mathbf{0}$ *(*$\boldsymbol{g} = \mathbf{0}$ *respectively) persists within* $\theta := \mathcal{O}(h^{2+\bar{\varkappa}})$ *(*$\theta := \mathcal{O}(h^{3+\bar{\varkappa}})$*),* $\bar{\varkappa} \in [0,1]$*, and the error estimates are then relaxed to an additional error term of size* $h^{-1}\theta$ *(*$h^{-2} \cdot \theta$*).*

Remark 5.16 (Eigenvalues of the propagation matrix)
For $n_{\boldsymbol{q}} = 1$*, the error amplification matrix in the above proof is given by*

$$\mathbf{T} = \begin{pmatrix} \dfrac{\alpha_m}{\alpha_m - 1} & 0 & 0 \\ 0 & \dfrac{\alpha_f}{\alpha_f - 1} & \dfrac{1 + \alpha_m - \gamma}{\gamma(\alpha_f - 1)} \\ 0 & 0 & \dfrac{\gamma - 1}{\gamma} \end{pmatrix}$$

if $\gamma \neq 0$ *and* $\alpha_m, \alpha_f \neq 1$ *such that the three eigenvalues are already given by the diagonal elements. The second order condition and* (5.13) *assure that* $\gamma = \frac{1}{2} - \alpha_m + \alpha_f > \frac{1}{2}$ *and so* $|\frac{\gamma-1}{\gamma}| < 1$*. By* $\alpha_m, \alpha_f < \frac{1}{2}$*, we also get* $|\frac{\alpha_*}{\alpha_* - 1}| < 1$*,* $* \in \{m, f\}$*, such that* $\varrho(\mathbf{T}) < 1$ *is proven. Note that for contractivity and so convergence, the condition* $\gamma > \frac{1}{2}$ *is already sufficient whereas* (4.3) *ensures second order convergence. With respect to the overshoot phenomenon (see Remark 4.21), it is also significant how the Jordan canonical form of the amplification matrix is structured. It can easily be verified that a degenerate Jordan decomposition (i. e., one with Jordan blocks of size* $m > 1$*) is present if and only if* $\alpha_f = 1 - \gamma$*. For the common settings of HHT, WBZ and CH(*ϱ_∞*) with* $\varrho_\infty \in [0, 1)$ *this is never the case.*

For Newmark integrators in the classical sense, it holds $\alpha_m = \alpha_f$ *such that* (5.13) *formally cannot be valid. Due to the equivalence of* $\boldsymbol{a}_n$ *and* $\dot{\boldsymbol{v}}_n$ *in these cases, the error analysis does not need to take into account* (4.2c) *and matrix* $\mathbf{T}$ *in the above analysis contains a redundant* *relation. However, second order convergence can only be gained for $\gamma = \frac{1}{2}$. Note that this always gives non-damping algorithms for (4.12).*

To understand why the expectably weaker numerical damping in Example 5.11 (a) for the settings $\varrho_\infty \in \{0.1716, 0.2679\}$ lead to numerically observed stronger damping, and why the presumably instantaneous annihilation for the choice $\varrho_\infty = 0$ could not be observed in the numerical test, we plug in the 'optimal parameters' of the Chung–Hulbert method according to (4.4). The eigenvalues of the above matrix $\mathbf{T}$ read

$$\mu_1 = \frac{\alpha_m}{\alpha_m - 1} = \frac{2\varrho_\infty - 1}{\varrho_\infty - 2}\,, \quad \mu_2 = \frac{\alpha_f}{\alpha_f - 1} = -\varrho_\infty\,, \quad \mu_3 = \frac{\gamma - 1}{\gamma} = \frac{3\varrho_\infty - 1}{\varrho_\infty - 3}$$

and so, the spectral radius of $\mathbf{T}$ is explicitly given by

$$\varrho(\mathbf{T}) = \begin{cases} \varrho_\infty & \text{if } \varrho_\infty \geq 2 - \sqrt{3} \\ \dfrac{2\varrho_\infty - 1}{\varrho_\infty - 2} & \text{if } 0 \leq \varrho_\infty < 2 - \sqrt{3}, \end{cases}$$

which indicates a maximal numerical damping for $\varrho_\infty = 2 - \sqrt{3} \approx 0.2679$. However, the displayed error in Figure 5.1 is basically $\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}$ whose (particularly transient) behavior is governed by powers of the lower two-by-two block, denoted by $\mathbf{T}_{2:3,2:3}$. As can straightforwardly be shown, see Figure 5.3, the spectral radius of $\mathbf{T}_{2:3,2:3}$ is minimized for $\varrho_\infty = 3 - 2\sqrt{2} \approx 0.1716$ which substantiates this specific choice.

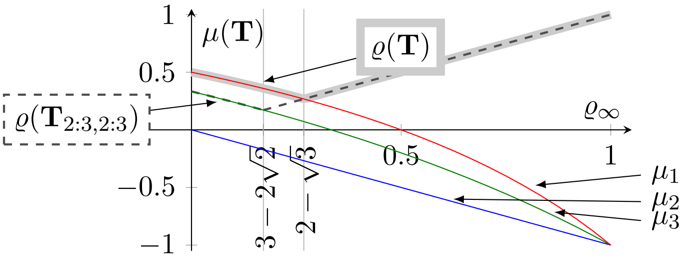


Figure 5.3: Eigenvalues of $\mathbf{T}$ for the CH(ϱ_∞) parameter set

Having a closer look at Figure 5.3, it also becomes evident that for $\varrho_\infty = 0$ no oscillations occur in the transient phase: Since all eigenvalues are nonnegative there is no switching of signs in each recursion. At last, note also that $\mathbf{T}$ is independent of the choice of β for this index-2 setting.

Remark 5.17 (Configuration-dependent order reduction and choice of $\boldsymbol{a}_0$)
It still remains to be shown why the order reduction in Example 5.11(b) only occurred for the initial configuration with $\dot{\boldsymbol{q}}(0) = (-1, 1)^\top$. The decoupling of the lowest order error terms in $\boldsymbol{E}_n^{\boldsymbol{z}}$ results in the conditions of Corollary 5.14 for the initial error terms on the level of acceleration-like variables. As a result, there is no order reduction if $\boldsymbol{e}_0^{\boldsymbol{a}} = \mathcal{O}(h)$ but $\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(h^2)$ (Arnold et al., 2016, Example 4.19). In general, this may be seen by Taylor expansion resulting in

$$\begin{aligned} \boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} &= \mathbf{G}(\boldsymbol{q}(t_0))\,(\ddot{\boldsymbol{q}}(t_0 + \Delta_\alpha h) - \boldsymbol{a}_0) = \mathbf{G}(\boldsymbol{q}(t_0))\,\big(\ddot{\boldsymbol{q}}(t_0) + \Delta_\alpha h \dddot{\boldsymbol{q}}(t_0) + \mathcal{O}(h^2) - \ddot{\boldsymbol{q}}(t_0)\big) \\ &= h\Delta_\alpha \mathbf{G}(\boldsymbol{q}(t_0))\dddot{\boldsymbol{q}}(t_0) + \mathcal{O}(h^2) \qquad \Rightarrow \quad \boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(h^2) \text{ if } \mathbf{G}(\boldsymbol{q}(t_0))\dddot{\boldsymbol{q}}(t_0) = \mathbf{0}\,. \end{aligned}$$

For the pendulum example with $\boldsymbol{q}(0) = \frac{1}{2}(\sqrt{2}, \sqrt{2})^\top$ it can be shown that

$$\mathbf{G}(\boldsymbol{q}(0))\dddot{\boldsymbol{q}}(0) = (\dot{q}_x(0) - \dot{q}_y(0))\,\Big(-g_{\text{grav}} + \tfrac{3}{4}\sqrt{2}\,\big((\dot{q}_x(0))^2 - (\dot{q}_y(0))^2\big)\Big)\,.$$

For zero velocity initial values we therefore get $\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(h^2)$ *and no order reduction in* $\dot{\boldsymbol{v}}$ *whereas for* $\dot{\boldsymbol{q}}(0) = (-1,1)^\top$ *we have* $\mathbf{G}(\boldsymbol{q}(0))\ddot{\boldsymbol{q}}(0) = 2g_{\text{grav}} \neq 0$ *and so a* $\mathcal{O}(h)$*-error in the transient phase.*

If one uses the more sophisticated initialization for the acceleration-like variables $\boldsymbol{a}_0$ *as presented in Section 4.2.3, it is also straightforward to show that the conditions of Corollary 5.14 are fulfilled and the order reduction can be avoided for arbitrary initial values* $(\boldsymbol{q}(t_0), \dot{\boldsymbol{q}}(t_0))^\top$*, see also the discussion in the index-3 case below.*

Remark 5.18 (Stabilized index-2 formulation)
As it is state-of-the-art (and quasi-standard) in technical simulation of multibody systems to use stabilization techniques for index-2 formulations, we will shortly discuss how the above analysis may also cover this setting. We briefly give an overview on this approach and its analysis as it is carried out in detail by Arnold et al. (2016).

The additional (velocity constraint enforcement) variable $\boldsymbol{\mu}_n \in \mathbb{R}^{n_{\boldsymbol{\lambda}}}$ *enters the algorithm by exchanging* (4.2a) *for*

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + h\boldsymbol{v}_n + h^2(\tfrac{1}{2} - \beta)\boldsymbol{a}_n + h^2\beta\boldsymbol{a}_{n+1} - h \cdot \mathbf{G}^\top(\boldsymbol{q}_n)\boldsymbol{\mu}_n$$

(Arnold, 2009) and, of course, consideration of both constraint equations in (5.7b)*. Note that the numerical solution for* $\boldsymbol{\mu}_n$ *enters only linearly and that* $\mathbf{G}^\top(\boldsymbol{q}_n)$ *is already known from the previous time step. The nonlinear systems in the algorithm nevertheless grow to dimension* $n_{\boldsymbol{q}} + 2n_{\boldsymbol{\lambda}}$*, taking into account the additional position constraints (Arnold and Hante, 2016). The local truncation errors* (5.4a) *for the position coordinates are then redefined and given by*

$$\begin{aligned}\boldsymbol{l}_n^{\boldsymbol{q},\text{GGL}} := \boldsymbol{q}(t_{n+1}) - \big(\boldsymbol{q}(t_n) + h\dot{\boldsymbol{q}}(t_n) + h^2(\tfrac{1}{2} - \beta)\ddot{\boldsymbol{q}}(t_n + \Delta_\alpha h)& \\ + h^2\beta\ddot{\boldsymbol{q}}(t_{n+1} + \Delta_\alpha h) - h\mathbf{G}(\boldsymbol{q}(t_n))\boldsymbol{\mu}(t_n)\big)&\,.\end{aligned}$$

Furthermore, the global error $\boldsymbol{e}_n^{\boldsymbol{\mu}} := \boldsymbol{\mu}_n$ *is formally introduced. Using Lemma 5.4, it can be estimated as*

$$\begin{aligned}\boldsymbol{e}_n^{\boldsymbol{\mu}} = -([\mathbf{G}\mathbf{G}^\top](\boldsymbol{q}(t_n)))^{-1}\mathbf{G}(\boldsymbol{q}(t_n))\,\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{q}}& \\ + \mathcal{O}(1)\Big(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\mu}}\|\Big) + \mathcal{O}(h^2)&\,,\end{aligned}$$

where we exploited the technical assumption on the global errors for $\boldsymbol{q}_n$ *again. Lemma 5.10 can now be used to show*

$$\|\boldsymbol{e}_n^{\boldsymbol{\mu}}\| = \mathcal{O}(1)\Big(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{e}_{n,n+1}^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\mu}}\|\Big) + \mathcal{O}(h^2)\,.$$

As a result, the errors in $\boldsymbol{\mu}_n$ *contribute only to the higher order terms if it is assumed that the corrector equations are solved such that* $\|\boldsymbol{g}(\boldsymbol{q}_n)\| = \mathcal{O}(h^3)$*. Finally, the results of Theorem 5.13 remain valid and the additional error bound* $\|\boldsymbol{e}_n^{\boldsymbol{\mu}}\| = \mathcal{O}(h^2)$ *holds.*

5.2.2 The index-3 case

As even in the index-2 case we have observed order reduction in the Lagrange multipliers $\boldsymbol{\lambda}_n$ for the Newmark integration family if the acceleration-like variables are not chosen carefully, it cannot be expected that for an implementation employing only position constraints, i. e., the equations of motion in its index-3 form, this issue is no longer present. One step of the algorithm in the index-3 setting is given by

$$\begin{aligned}
\boldsymbol{q}_{n+1} &= \boldsymbol{q}_n + h\boldsymbol{v}_n + h^2(\tfrac{1}{2}-\beta)\boldsymbol{a}_n + h^2\beta\boldsymbol{a}_{n+1}\,,\\
\boldsymbol{v}_{n+1} &= \boldsymbol{v}_n + h(1-\gamma)\boldsymbol{a}_n + h\gamma\boldsymbol{a}_{n+1}\,,\\
(1-\alpha_m)\boldsymbol{a}_{n+1} + \alpha_m\boldsymbol{a}_n &= (1-\alpha_f)\dot{\boldsymbol{v}}_{n+1} + \alpha_f\dot{\boldsymbol{v}}_n\,,\\
\mathbf{M}(\boldsymbol{q}_{n+1})\dot{\boldsymbol{v}}_{n+1} &= \boldsymbol{f}(\boldsymbol{q}_{n+1},\boldsymbol{v}_{n+1}) - \mathbf{G}^\top(\boldsymbol{q}_{n+1})\boldsymbol{\lambda}_{n+1}\\
\boldsymbol{g}(\boldsymbol{q}_{n+1}) &= \mathbf{0}
\end{aligned}$$

with $\boldsymbol{q}_0 = \boldsymbol{q}(t_0)$. We will see in (5.17) and Remark 5.23 below, that it may be advantageous to use a different initialization for the velocity variables $\boldsymbol{v}_0$ than the obvious choice $\dot{\boldsymbol{q}}(t_0)$ but if not explicitly stated otherwise we will still use the original setting, especially $\dot{\boldsymbol{v}}_0 = \boldsymbol{a}_0 = \ddot{\boldsymbol{q}}(t_0)$.

The analysis of the method in the index-3 case, which is mainly adapted from (Arnold et al., 2015a), will essentially use the same results as the one in the previous section but the condensed error vectors $\boldsymbol{E}_n^{\boldsymbol{z}}$ will take one more component into account such that its propagation behavior in the end is much closer to the one from the harmonic oscillator example from Chapter 4 and it also allows for an easy way to construct remedies for the inferior convergence behavior to be described in the following instructive example.

Example 5.19 (Arnold et al. (2015a, Example 2))
We consider the scalar (and pathological) test example

$$\ddot{q}(t) = -\lambda(t)\,, \tag{5.16a}$$

$$q(t) - t^3 = 0\,, \quad (q(0) = \dot{q}(0) = 0,\ t \in [0,T])\,. \tag{5.16b}$$

The main difficulties that arise for Newmark integrators in an index-3 setting are revealed for this simple polynomial test problem already. In the above form, the constraint $g = g(t,q) = q - t^3$ is *rheonomic* and $n_{\boldsymbol{\lambda}} = n_{\boldsymbol{q}}$, i. e., the problem does not exactly resemble the proposed structure of the mechanical systems under consideration but simplifies the reasoning. The problem obeys the analytic solution

$$q(t) = t^3\,, \quad \dot{q}(t) = 3t^2\,, \quad \ddot{q}(t) = -\lambda(t) = 6t\,.$$

Consistent initialization of the acceleration variables is given by $\dot{v}_0 := 0$, but for this example we will treat $\dot{v}_0$ as a free variable. The first integration step of the Newmark integrator comprises the system

$$\begin{aligned}
q_1 &= h^2(\tfrac{1}{2}-\beta)a_0 + h^2\beta a_1\,,\\
v_1 &= h(1-\gamma)a_0 + h\gamma a_1\,,\\
(1-\alpha_m)a_1 + \alpha_m a_0 &= (1-\alpha_f)\dot{v}_1 + \alpha_f\dot{v}_0\,,\\
q_1 &= h^3\\
\dot{v}_1 &= -\lambda_1
\end{aligned}$$

which for $\beta \neq 0$, $\alpha_f \neq 1$ has the explicit solution

$$\begin{aligned}
q_1 &= h^3\,,\\
v_1 &= \frac{(2\beta-\gamma)a_0}{2\beta}h + \frac{\gamma}{\beta}h^2\,,\\
\dot{v}_1 = -\lambda_1 &= \frac{(2\beta-1+\alpha_m)a_0 - 2\beta\alpha_f\dot{v}_0}{2(1-\alpha_f)\beta} + \frac{1-\alpha_m}{(1-\alpha_f)\beta}h\,.
\end{aligned}$$

A comparison with the exact solutions at $t = h$ shows the error behavior

$$\begin{aligned}
e_1^{\boldsymbol{q}} &= 0\,,\\
e_1^{\boldsymbol{v}} &= \mathcal{O}(h^2) \text{ if } a_0 = \mathcal{O}(h),\\
e_1^{\boldsymbol{\lambda}} &= \frac{(1-\alpha_m-2\beta)a_0 + 2\alpha_f\beta\dot{v}_0}{2(1-\alpha_f)\beta} + \left(6 - \frac{1-\alpha_m}{(1-\alpha_f)\beta}\right)h\,.
\end{aligned}$$

Whatever choices (that do not explicitly depend on the time step size) for the variables a_0 and even $\dot{v}_0$, are made, it is not possible to avoid that the order of the approximations on Lagrange multiplier (and also acceleration) level is just one. Due to the good numerical damping properties of the algorithms, this parasitic error components are damped out for strictly stable methods after a transient phase as will be shown below. ♢

The order reduction we have seen in the previous example occurs whenever the Newmark integrator needs to be initialized. For models with contact conditions or discontinuous right hand sides as well as for systems with control inputs this may happen many times during the time integration process. When a variable time step scheme is used, the order reduction may even be observed whenever a change in the step size occurs (see Arnold et al., 2015b). To remedy this unsatisfactory property, Arnold et al. (2015a) propose to add corrector terms

$$\boldsymbol{v}_0 := \dot{\boldsymbol{q}}(t_0) + \boldsymbol{\delta}^{\boldsymbol{v}}_{\text{corr}} \quad \text{and} \quad \boldsymbol{a}_0 := \ddot{\boldsymbol{q}}(t_0) + \boldsymbol{\delta}^{\boldsymbol{a}}_{\text{corr}}$$

$$\begin{aligned}
\text{with} \quad \boldsymbol{\delta}^{\boldsymbol{v}}_{\text{corr}} &:= (1-6\beta+3\Delta_\alpha)\frac{h^2}{6}\left[\mathbf{M}^{-1}\mathbf{G}^\top\mathbf{S}^{-1}\mathbf{G}\right](\boldsymbol{q}_0)\frac{\dot{\boldsymbol{v}}_{sh}-\dot{\boldsymbol{v}}_{-sh}}{2sh},\\
\text{and} \quad \boldsymbol{\delta}^{\boldsymbol{a}}_{\text{corr}} &:= \Delta_\alpha h\frac{\dot{\boldsymbol{v}}_{sh}-\dot{\boldsymbol{v}}_{-sh}}{2sh}\,,
\end{aligned} \tag{5.17}$$

to the starting values of the the velocity and acceleration-like coordinates to preserve second order. The terms $\dot{\boldsymbol{v}}_{\pm sh}$ denote approximations to $\ddot{\boldsymbol{q}}$ at $t = t_0 \pm sh$ with a (small) constant $s \in (0,1)$, such that $\frac{\dot{\boldsymbol{v}}_{sh}-\dot{\boldsymbol{v}}_{-sh}}{2sh}$ approximates the third derivative of $\boldsymbol{q}$. From Corollary 5.14 and the above error analysis, the definition of $\boldsymbol{\delta}^{\boldsymbol{a}}_{\text{corr}}$ is evident: Perturbing $\boldsymbol{a}_0$ by adding a term of magnitude $\mathcal{O}(h)$ has no effect on the initial error estimates from Lemma 5.12 but with that choice the order of $\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}$ is automatically two instead of one, cf. Remark 5.17. The particular choice of $\boldsymbol{\delta}^{\boldsymbol{v}}_{\text{corr}}$ will become clearer in the error analysis below.

Lemma 5.20 (Initial error terms (II))
For the index-3 Newmark integrator with starting values

$$\boldsymbol{v}_0 = \dot{\boldsymbol{q}}(t_0)\,, \quad \boldsymbol{a}_0 = \dot{\boldsymbol{v}}_0 = \ddot{\boldsymbol{q}}(t_0)$$

the assertions on the initial error terms on level of acceleration-like variables from Lemma 5.12 remain valid.

Moreover, it holds

$$\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}} + \tfrac{1}{h}\boldsymbol{l}_0^{\mathbf{G}\boldsymbol{q}}\| = \mathcal{O}(h^2)\,. \tag{5.18}$$

Proof. The assertion is a simple corollary of (5.4b) and the fact that for the above initialization $\boldsymbol{v}_0 := \dot{\boldsymbol{q}}(t_0)$ the initial error term $\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}}$ vanishes. Note, however, that estimate (5.18) can be improved with (5.4b) if $\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}} + \frac{h^2}{6}(1-6\beta+3(\alpha_m-\alpha_f))\mathbf{G}(\boldsymbol{q}_0)\dddot{\boldsymbol{q}}(t_0)\|$ is sufficiently small. □

We pointed out in the outline of this section that for the convergence analysis in the index-3 case the error vector $\boldsymbol{E}_n^{\boldsymbol{z}}$ will be extended by another component. To this end, we introduce the (curvature) error term

$$\boldsymbol{r}_n^{\mathbf{G}} := \frac{1}{h}\left(\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} + \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_n^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_n)) + \tfrac{1}{h}\boldsymbol{l}_n^{\mathbf{G}\boldsymbol{q}}\right)\,,\ n = 0, 1, \dots\,,$$

and the notation

$$\boldsymbol{\eta}_n^{(3)} := \|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\|\,,\ n = 0, 1, \dots$$

to collect higher order terms in the following estimates, because for this and the SPP setting in the sections to follow below, the higher order terms have a more complicated structure. The superscript $\bullet^{(3)}$ just indicates the index-3 case.

Theorem 5.21 (Convergence in the index-3 DAE case, Arnold et al. (2015a, Theorem 1)) Let a stable second order Newmark integrator (4.2) in the sense of Definition 4.16 be given. Suppose that the corrector equations are solved such that $\max_m\|\boldsymbol{g}(\boldsymbol{q}_m)\| = \mathcal{O}(h^4)$.

(a) If the starting values satisfy

$$\|\boldsymbol{e}_0^{\boldsymbol{q}}\| + \|\boldsymbol{e}_0^{\boldsymbol{v}}\| = \mathcal{O}(h^2)\,, \quad \|\boldsymbol{e}_0^{\boldsymbol{a}}\| + \|\boldsymbol{e}_0^{\boldsymbol{\lambda}}\| = \mathcal{O}(h)$$

then for the errors of the Newmark integrator in the index-3 setting the relation

$$\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| \le C\,\mathrm{e}^{\tilde{L}(t_n - t_0)}\,h^2\,,$$

$$\left\|\begin{pmatrix}\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}\\ \boldsymbol{r}_n^{\mathbf{G}}\\ \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}\\ \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\end{pmatrix} - \mathbf{T}^n\begin{pmatrix}\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}}\\ \boldsymbol{r}_0^{\mathbf{G}}\\ \boldsymbol{e}_0^{\mathbf{S}\boldsymbol{\lambda}}\\ \boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}\end{pmatrix}\right\| \le C\,\mathrm{e}^{\tilde{L}(t_n - t_0)}\,h^2$$

holds where $\mathbf{T} := \mathrm{blkdiag}(-\alpha_m/(1-\alpha_m)\mathbf{I}_{n_q}, \mathbf{T}(\infty)\otimes\mathbf{I}_{n_{\boldsymbol{\lambda}}})$ with $\mathbf{T}(\infty) := \lim_{z\to\infty}\mathbf{T}(z)$ denoting the error amplification matrix for the linear test equation (4.12) in the limit case $h\omega \to \infty$, cf. (4.16).

(b) If the starting values additionally satisfy

$$\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}} + \mathsf{R}(\boldsymbol{q}(t_0))(\boldsymbol{e}_0^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_0)) + \tfrac{1}{h}\boldsymbol{l}_0^{\mathbf{G}\boldsymbol{q}}\| + h\|\boldsymbol{e}_0^{\boldsymbol{\lambda}}\| + h\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}\| = \mathcal{O}(h^{2+\varkappa})\,, \quad \varkappa \in [0,1]\,,$$

then the global errors on Lagrange multiplier level of the Newmark integrator in its index-3 form are bounded like

$$\|\boldsymbol{e}_n^{\boldsymbol{\lambda}}\| \le C_0\left(\varrho^n h^{1+\varkappa} + \mathrm{e}^{\tilde{L}(t_n - t_0)}\,h^2\right)\,,$$

where $0 < \varrho < 1$ is a constant depending on the parameters of the algorithm.

The positive constants C_0, $\tilde{L}$ are independent of n.

Proof. By definition of $\boldsymbol{r}_n^{\mathbf{G}}$ and (5.6b) in Corollary 5.5, one can directly show that

$$\boldsymbol{r}_n^{\mathbf{G}} + (\tfrac{1}{2} - \beta)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + \beta\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} = \tfrac{1}{h}(\mathbf{G}(\boldsymbol{q}(t_n))\,\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{q}} + \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_n^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_n))) + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{a}}\|\,.$$

Lemma 5.9 then leads to

$$\boldsymbol{r}_n^{\mathbf{G}} + (\tfrac{1}{2} - \beta)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + \beta\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(3)} + \mathcal{O}(h^2) - \tfrac{1}{h}\,\boldsymbol{\Delta}_h\,\boldsymbol{g}(\boldsymbol{q}_n)\,. \tag{5.19}$$

This relation will replace equation (5.15) from the error recursion formula of the index-2 setting. For the index-3 case another relation is obtained considering the difference of two instances $\boldsymbol{r}_\bullet^{\mathbf{G}}$:

$$\begin{aligned}
\boldsymbol{r}_{n+1}^{\mathbf{G}} - \boldsymbol{r}_n^{\mathbf{G}} &= \boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} + \mathcal{O}(1)(\|\boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_{n+1}^{\boldsymbol{q}}\|) + \tfrac{1}{h^2}(\boldsymbol{l}_{n+1}^{\mathbf{G}\boldsymbol{q}} - \boldsymbol{l}_n^{\mathbf{G}\boldsymbol{q}}) \\
&= (1-\gamma)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + \gamma \boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \tfrac{1}{h}\mathbf{G}(\boldsymbol{q}(t_n))\, \boldsymbol{\Delta}_h\, \boldsymbol{l}_n^{\boldsymbol{q}} + \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(3)} + \mathcal{O}(h^2) + \underbrace{\mathcal{O}(1)\tfrac{1}{h}\boldsymbol{l}_{n+1}^{\boldsymbol{q}}}_{=\mathcal{O}(h^2)} \\
&= (1-\gamma)\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + \gamma \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} + \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(3)} + \mathcal{O}(h^2)\,, \qquad (5.20)
\end{aligned}$$

where, at first, we used

$$\begin{aligned}
&\mathsf{R}(\boldsymbol{q}(t_{n+1}))(\boldsymbol{e}_{n+1}^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_{n+1})) - \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_n^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_n)) \\
&\qquad = \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_{n+1}^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_{n+1})) + \mathcal{O}(h)\|\boldsymbol{e}_{n+1}^{\boldsymbol{q}}\| - \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_{n+1}^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_n)) \\
&\qquad\quad + \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_{n+1}^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_n)) - \mathsf{R}(\boldsymbol{q}(t_n))(\boldsymbol{e}_n^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_n)) \\
&\qquad = \mathcal{O}(h)\left(\|\boldsymbol{e}_{n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{\Delta}_h\, \boldsymbol{e}_n^{\boldsymbol{q}}\|\right)
\end{aligned}$$

and then (5.6d) and (5.4). The finite differences of $\boldsymbol{e}_n^{\boldsymbol{q}}$ were estimated with the aid of (5.6a). The findings from Lemma 5.8 remain valid in the index-3 case as well and—in terms of $\boldsymbol{\eta}_n^{(3)}$—may be expressed as

$$\begin{aligned}
&(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} = \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(3)} + \mathcal{O}(h^2)\,, \\
&(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} = -(1-\alpha_f)\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda}} - \alpha_f \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}} + \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(3)} + \mathcal{O}(h^2)\,, \qquad (5.21)
\end{aligned}$$

where $\mathbf{S} = \mathbf{S}(\boldsymbol{q})$ denotes again the regular Delassus matrix from (2.20).

The new condensed error terms are $\boldsymbol{E}_n^{\boldsymbol{y}} := (\boldsymbol{e}_n^{\boldsymbol{q}}, \boldsymbol{e}_n^{\boldsymbol{v}})^\top$, $\boldsymbol{E}_n^{\boldsymbol{z}} := (\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}, \boldsymbol{r}_n^{\mathbf{G}}, \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}, \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}})^\top$, where for future reference the last three components are denoted by $\boldsymbol{E}_n^{\boldsymbol{r}} := (\boldsymbol{r}_n^{\mathbf{G}}, \boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda}}, \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}})^\top$. The first set of inequalities in Lemma 5.1 remains unchanged

$$\boldsymbol{E}_{n+1}^{\boldsymbol{y}} = \boldsymbol{E}_n^{\boldsymbol{y}} + \mathcal{O}(h)(\|\boldsymbol{E}_n^{\boldsymbol{y}}\| + \|\boldsymbol{E}_{n,n+1}^{\boldsymbol{z}}\|) + \mathcal{O}(h^2)$$

and all the above estimates can again be expressed within the framework of Lemma 5.1 with

$$\mathbf{T} := \operatorname{blkdiag}\left(\frac{-\alpha_m \mathbf{I}_{n_q}}{1-\alpha_m}, \underbrace{((\tilde{\mathbf{T}}_1^{(3)})^{-1} \cdot \tilde{\mathbf{T}}_2^{(3)})}_{\mathbf{T}(\infty)} \otimes \mathbf{I}_{n_{\boldsymbol{\lambda}}})\right), \quad M = \mathcal{O}(h^2)\,,$$

where the matrices $\tilde{\mathbf{T}}_{1|2}^{(3)}$ are known from the infinite stiffness case of the linear stability analysis in Chapter 4, i. e.,

$$\tilde{\mathbf{T}}_1^{(3)} := \begin{pmatrix} 0 & 0 & -\beta \\ 1 & 0 & -\gamma \\ 0 & 1-\alpha_f & 1-\alpha_m \end{pmatrix}, \quad \tilde{\mathbf{T}}_2^{(3)} := \begin{pmatrix} 1 & 0 & \frac{1}{2}-\beta \\ 1 & 0 & 1-\gamma \\ 0 & -\alpha_f & -\alpha_m \end{pmatrix}, \qquad (5.22)$$

see (4.17) which gives an amplification matrix $\mathbf{T}$ as given in the theorem. More precisely, (5.19), (5.20) and (5.21) can be combined using the assumption on the constraint residuals $\boldsymbol{g}(\boldsymbol{q}_n) = \mathcal{O}(h^4) \Rightarrow \frac{1}{h}\boldsymbol{\Delta}_h\, \boldsymbol{g}(q_n) = \mathcal{O}(h^2)$ to get

$$(\tilde{\mathbf{T}}_1^{(3)} \otimes \mathbf{I}_{n_{\boldsymbol{\lambda}}})\boldsymbol{E}_{n+1}^{\boldsymbol{r}} - (\tilde{\mathbf{T}}_2^{(3)} \otimes \mathbf{I}_{n_{\boldsymbol{\lambda}}})\boldsymbol{E}_n^{\boldsymbol{r}} = \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(3)} + \mathcal{O}(h^2)\,, \qquad (5.23)$$

such that the result follows as for Theorem 5.13. Note that the contribution of $\mathsf{R}(\boldsymbol{q}(t_0))(\boldsymbol{e}_0^{\boldsymbol{q}}, \dot{\boldsymbol{q}}(t_0))$ to the initial error term $\boldsymbol{r}_0^{\mathbf{G}}$ only regards higher order terms since it is linear in $\boldsymbol{e}_0^{\boldsymbol{q}}$. Note also that the conditions for well-definition of $\mathbf{T}$, i. e., $\alpha_f, \alpha_m \neq 1$, $\beta \neq 0$ are automatically fulfilled since we required the method to be stable. □

Corollary 5.22 (Order reduction depending on initial data)
From Theorem 5.21 it is evident that for the initialization $\boldsymbol{q}_0 = \boldsymbol{q}(t_0)$ a first order transient error term for the Lagrange multipliers is always present if only (5.18) holds. The Newmark integrator therefore preserves its second order of convergence, even for the Lagrange multipliers, if

$$\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}} + \tfrac{1}{h}\boldsymbol{l}_0^{\mathbf{G}\boldsymbol{q}}\| + h\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}\| = \mathcal{O}(h^3)\,.$$

Remark 5.23 (Proof of convergence and adapted initial values)
If we were only to prove first order convergence of the Lagrange multipliers, the analysis could have been substantially simplified. Especially, there would not have been the need for introducing $\boldsymbol{r}_n^{\mathbf{G}}$*, the coupled error propagation would resemble the one we are about to give in Section 5.3.2 below.*

Employing the complete error analysis we are now able to justify the correction terms $\boldsymbol{\delta}_{\mathrm{corr}}^{\boldsymbol{v}}$ *and* $\boldsymbol{\delta}_{\mathrm{corr}}^{\boldsymbol{a}}$ *in (5.17) from the introduction of this section. Corollary 5.22 indicates that order reduction of the index-3 integrator can always be avoided if we use the correction term* $\boldsymbol{\delta}_{\mathrm{corr}}^{\boldsymbol{a}}$ *to get* $\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(h^2)$ *as in Section 4.2.3 and so that Theorem 5.21 guarantees second order of convergence if* $\mathbf{G}(\boldsymbol{q}(t_0))(\boldsymbol{e}_0^{\boldsymbol{v}} + \frac{1}{h}\boldsymbol{l}_0^{\boldsymbol{q}}) = \mathcal{O}(h^3)$*, which is satisfied if the correction term* $\boldsymbol{\delta}_{\mathrm{corr}}^{\boldsymbol{v}}$ *fulfills*

$$\mathbf{G}(\boldsymbol{q}_0)(\dot{\boldsymbol{q}}(t_0) - \boldsymbol{v}_0) = \mathbf{G}(\boldsymbol{q}_0)\boldsymbol{\delta}_{\mathrm{corr}}^{\boldsymbol{v}} = \tfrac{1}{h}\mathbf{G}(\boldsymbol{q}_0)\boldsymbol{l}_0^{\boldsymbol{q}} \overset{\bullet}{=} \frac{1 - 6\beta + 3\Delta_\alpha}{6}\mathbf{G}(\boldsymbol{q}_0)\dddot{\boldsymbol{q}}(t_0)\,, \tag{5.24}$$

where "$\overset{\bullet}{=}$*" indicates identity up to higher order terms. Since this system is underdetermined, one may add the equation* $\mathbf{M}(\boldsymbol{q}_0)\boldsymbol{\delta}_{\mathrm{corr}}^{\boldsymbol{v}} + \mathbf{G}^\top(\boldsymbol{q}_0)\boldsymbol{\delta}_{\mathrm{corr}}^{\mathrm{dummy}} = \mathbf{0}$ *with the additional variable* $\boldsymbol{\delta}_{\mathrm{corr}}^{\mathrm{dummy}} \in \mathbb{R}^{n_{\boldsymbol{\lambda}}}$*. With (5.24) the correction terms may thus be uniquely defined by solving the linear system*

$$\begin{pmatrix} \mathbf{M}(\boldsymbol{q}_0) & \mathbf{G}^\top(\boldsymbol{q}_0) \\ \mathbf{G}(\boldsymbol{q}_0) & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{\mathrm{corr}}^{\boldsymbol{v}} \\ \boldsymbol{\delta}_{\mathrm{corr}}^{\mathrm{dummy}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \dfrac{1 - 6\beta + 3\Delta_\alpha}{6}\mathbf{G}(\boldsymbol{q}_0)\dddot{\boldsymbol{q}}(t_0) \end{pmatrix},$$

where $\dddot{\boldsymbol{q}}(t_0)$ *can be approximated by means of finite differences. This definition of the velocity correction is convenient because it is very cheap to compute as the involved saddle-point matrix needs to be decomposed for the determination of initial values (and possibly within the time integration itself) anyway.*

Note that the correction term for the velocity components is in $\mathcal{O}(h^2)$ *such that the convergence result for* $(\boldsymbol{q}_n, \boldsymbol{v}_n)^\top$ *remains valid and that* $\mathbf{M}(\boldsymbol{q}_0) = \mathbf{M}(\boldsymbol{q}(t_0 + \Delta_\alpha h)) + \mathcal{O}(h)$*,* $\mathbf{G}(\boldsymbol{q}_0) = \mathbf{G}(\boldsymbol{q}(t_0 + \Delta_\alpha h)) + \mathcal{O}(h)$ *and so, it is justified to take these arguments.*

It is also important that, see Remark 5.17, $\mathbf{G}(\boldsymbol{q}(t_0))\dddot{\boldsymbol{q}}(t_0) = \mathbf{0}$ *is sufficient for preventing order reduction in both, the index-2 and the index-3 case.*

Remark 5.24 (Exact fulfillment of velocity constraints (Arnold et al., 2016, Lemma 3.4 b)))
With regard to the important application field of structural dynamics where the numerical challenge often lies in the large dimension of the system rather than the complicated nonlinear arrangement of force vector and constraints (Simeon, 2013), the case of linear constraints *plays an important role and shall be mentioned here in a little more detail.*

For constant constraint Jacobian $\mathbf{G}(\boldsymbol{q}) \equiv \mathbf{G}$ *and consistent initial values*

$$\boldsymbol{g}(\boldsymbol{q}_0) = \mathbf{G}\boldsymbol{q}_0 + \boldsymbol{g}_0 = \mathbf{G}\boldsymbol{v}_0 = \mathbf{G}\boldsymbol{a}_0 = \mathbf{0}\,, \quad \boldsymbol{g}_0 \in \mathbb{R}^{n_{\boldsymbol{\lambda}}}\,,$$

the index-3 integrator fulfills the hidden constraints on velocity level exactly. The proof is an elementary induction over the time steps. As a result, the convergence result of the index-2 case can be carried over such that neither order reduction nor large overshoot is present. Note that this observation has important implications for the model setup: As the convergence behavior of *the algorithm is substantially improved in case of linear constraints, the choice of coordinates may influence the performance of the simulation. For configuration spaces with Lie-group structure a comprehensive discussion is presented by Müller and Terze (2014). Also, in case of more general multibody systems it is sometimes appropriate to consider (easier) linearized equations such that also the computation of constraint- and reaction forces and the setup of variables for optimal control is simplified (Eich-Soellner and Führer, 1998).*

In case of linear constraint equations (corresponding to a quadratic penalty potential for the SPPs) the analysis can be simplified to a (more or less plain) extension of the harmonic oscillator example from Section 4.2. Note, nonetheless, that the analysis itself is governed by the linearized equations of motion leading to the coupled onestep error propagation in (5.23) *resembling the analysis of the linear ODE case. Incidentally, for Runge–Kutta methods it is well-known that for certain Hessenberg systems (see Example 2.9) with linear constraints, the hidden constraints are automatically preserved as well (see e. g. Nipp, 2002) and even give the same numerical results when applied to the systems in its different formulations.*

5.3 Singularly perturbed systems

To study the numerical properties of (4.2) in the SPP case, we will carry over as much as possible from the DAE analysis of the previous section. In particular, this means that all estimates that do not include any Lagrange multiplier terms remain unchanged since the error terms are defined with respect to the slow solution $\boldsymbol{q}(t)$ of the DAE systems anyway, cf. (5.3c) and (5.3d).

Initial values that are consistent with the according differential-algebraic systems are an unrealistic assumption but with respect to the singular nature of the perturbation force terms in (3.6) and (3.15), it seems reasonable to at least presume some upper bound on the initial energy of the system. For the initial deviations from the constraint manifolds of the slow system, this implies the following conditions.

Assumption 5.25 (Initial deviations)**.**
The initial values of the SPPs fulfill the following estimates

(a) $\boldsymbol{g}(\boldsymbol{q}^\delta(t_0)) = \mathcal{O}(h^2)$, $\mathbf{G}(\boldsymbol{q}^\delta(t_0))\dot{\boldsymbol{q}}^\delta(t_0) = \mathcal{O}(\delta)$ in the strongly damped case and

(b) $\boldsymbol{g}(\boldsymbol{q}^\varepsilon(t_0)) = \mathcal{O}(\varepsilon^2)$, $\mathbf{G}(\boldsymbol{q}^\varepsilon(t_0))\dot{\boldsymbol{q}}^\varepsilon(t_0) = \mathcal{O}(\varepsilon^2)$ for stiff mechanical systems.

Here, $h > 0$ denotes the time step size of (4.2) which is assumed to be bounded from above by a sufficiently small constant $h_0 > 0$.

Note that these assumptions are stronger than the ones imposed for the convergence results in case of Runge–Kutta methods in Remarks 3.11 and 3.20. In both cases the initial error terms were allowed to have a deviation that depends on the time step size $h > \delta$ ($h > \varepsilon$ respectively) which in view of Assumption 5.3 is in fact a much weaker assumption. The numerical results from Chapters 1 and 6 demonstrate that in this case convergence can no longer be guaranteed. The h^2 bound on position level for the strongly damped case is motivated by the classical second order of the method which we want to preserve in the SPP case if possible. Note, however, that this deviation may cause a larger drift from the constraint manifold $\mathfrak{M}^{\mathrm{s}}$. Because h is bounded, the unique definition of the corresponding slow motion is always possible.

As done in Assumption 5.6, we will impose a technical assumption on the error terms in the SPP case also. The singular nature of the problems forbids to give (global) convergence results on acceleration level for general initial values as they are declared in Assumption 5.25. So, we will only include the position and velocity variables, which also makes sense if we only consider the SPPs as substitute problems to DAEs: In that case, the singular forces in the differential equations have no physical meaning anyway. As for Assumption 5.6 in the DAE case, Assumption 5.26 below can be justified by induction afterwards because the final error estimates will provide stronger bounds.

Assumption 5.26 (Technical assumption, the SPP case).
For the analysis in the SPP case we will suppose that there exist constants $C, h_0 > 0$, (independent of the penalty parameters δ and ε), such that whenever $0 < h \leq h_0$ holds, we have the estimates

$$\begin{array}{lll} \|\boldsymbol{e}_m^{\boldsymbol{q}}\| < Ch\,, & \|\boldsymbol{e}_m^{\boldsymbol{v}}\| < Ch & \text{for strongly damped systems,} \\ \|\boldsymbol{e}_m^{\boldsymbol{q}}\| < Ch\,, & \|\boldsymbol{e}_m^{\boldsymbol{v}}\| < C & \text{for stiff mechanical systems} \end{array}$$

for all $m \geq 0$, $t_0 + mh \leq t_{\text{end}}$.

5.3.1 Strongly damped systems

When applied to strongly damped mechanical systems (3.6), (3.7) respectively, algorithm (4.2) reads

$$\begin{array}{rrcl} & \boldsymbol{q}_{n+1}^{\delta} & = & \boldsymbol{q}_n^{\delta} + h\boldsymbol{v}_n^{\delta} + h^2(\tfrac{1}{2} - \beta)\boldsymbol{a}_n^{\delta} + h^2\beta\boldsymbol{a}_{n+1}^{\delta}\,, \\ & \boldsymbol{v}_{n+1}^{\delta} & = & \boldsymbol{v}_n^{\delta} + h(1-\gamma)\boldsymbol{a}_n^{\delta} + h\gamma\boldsymbol{a}_{n+1}^{\delta}\,, \\ & (1-\alpha_m)\boldsymbol{a}_{n+1}^{\delta} + \alpha_m\boldsymbol{a}_n^{\delta} & = & (1-\alpha_f)\dot{\boldsymbol{v}}_{n+1}^{\delta} + \alpha_f\dot{\boldsymbol{v}}_n^{\delta}\,, \\ \text{a)}\ \Big\{ & \mathbf{M}(\boldsymbol{q}_{n+1}^{\delta})\dot{\boldsymbol{v}}_{n+1}^{\delta} & = & \boldsymbol{f}(\boldsymbol{q}_{n+1}^{\delta}, \boldsymbol{v}_{n+1}^{\delta}) - \tfrac{1}{\delta}[\mathbf{G}^{\top}\mathbf{G}](\boldsymbol{q}_{n+1}^{\delta})\boldsymbol{v}_{n+1}^{\delta}\,, \\ \text{or b)}\ \left\{\begin{array}{l} \\ \\ \end{array}\right. & \begin{array}{r} \mathbf{M}(\boldsymbol{q}_{n+1}^{\delta})\dot{\boldsymbol{v}}_{n+1}^{\delta} \\ \delta\boldsymbol{\lambda}_{n+1}^{\delta} \end{array} & \begin{array}{c} = \\ = \end{array} & \begin{array}{l} \boldsymbol{f}(\boldsymbol{q}_{n+1}^{\delta}, \boldsymbol{v}_{n+1}^{\delta}) - \mathbf{G}^{\top}(\boldsymbol{q}_{n+1}^{\delta})\boldsymbol{\lambda}_{n+1}^{\delta}\,, \\ \mathbf{G}(\boldsymbol{q}_{n+1}^{\delta})\boldsymbol{v}_{n+1}^{\delta}\,. \end{array} \end{array} \tag{5.25}$$

Note that the second formulation b) can only be used in a practical implementation when the involved functions $\mathbf{M}$, $\boldsymbol{f}$ and $\mathbf{G}$ can explicitly evaluated which often is not the case. Yet, for the error analysis it will prove useful to rely on this formulation although the Lagrange multiplier variables $\boldsymbol{\lambda}_n^{\delta}$ are not calculated in that case.

Before the comprehensive error study from the previous section is adapted to SPPs, we consider again two simple examples to provide a first insight to similarities and differences.

Example 5.27 (Linear analysis: The attractive equivalent of the harmonic oscillator)
A straightforward index reduction for the constrained equivalent of the harmonic oscillator, cf. (4.12), leads to the singularly perturbed problem

$$\ddot{q}^{\delta}(t) + \frac{1}{\delta}\dot{q}^{\delta}(t) = 0\,, \tag{5.26}$$

which can analytically be solved for the solution

$$q^{\delta}(t) = q^{\delta}(t_0) + \delta\dot{q}^{\delta}(t_0)\left(1 - \mathrm{e}^{-t/\delta}\right)\,.$$

A numerical integration method can no longer be required to damp out the solutions entirely for this would not take into account that all analytic solutions comprise a (possibly non-vanishing) constant component. Nevertheless, solutions of (5.26) still remain bounded on arbitrary time intervals and this should be reflected for approximate solutions as well. The onestep recursion corresponding to (4.16) for the harmonic oscillator now reads

$$\begin{pmatrix} 1 & 0 & -\beta \\ 0 & \delta/h & -\gamma \\ 0 & 1-\alpha_f & 1-\alpha_m \end{pmatrix} \begin{pmatrix} q_{n+1}^{\delta} \\ h^2\delta^{-1}v_{n+1}^{\delta} \\ h^2a_{n+1}^{\delta} \end{pmatrix} = \begin{pmatrix} 1 & \delta/h & \frac{1}{2}-\beta \\ 0 & \delta/h & 1-\gamma \\ 0 & -\alpha_f & -\alpha_m \end{pmatrix} \begin{pmatrix} q_n^{\delta} \\ h^2\delta^{-1}v_n^{\delta} \\ h^2a_n^{\delta} \end{pmatrix}\,.$$

The existence of constant solution components is reflected by the invariant subspace of this linear mapping $(q_n^\delta, h^2\delta^{-1}v_n^\delta, h^2 a_n^\delta) \mapsto (q_{n+1}^\delta, h^2\delta^{-1}v_{n+1}^\delta, h^2 a_{n+1}^\delta)$ to the eigenvector $(1,0,0)^\top$ with an eigenvalue that is exactly $\mu_1 = 1$. So, the damping properties of the Newmark integrators are completely characterized by the amplification behavior of the lower two-by-two block:

$$\mathbf{T}_\delta := \begin{pmatrix} \delta/h & -\gamma \\ 1-\alpha_f & 1-\alpha_m \end{pmatrix}^{-1} \begin{pmatrix} \delta/h & 1-\gamma \\ -\alpha_f & -\alpha_m \end{pmatrix},$$

which for $\delta = 0$ coincides with (the lower block of) the error propagation of the index-2 case. We notice that even for $\delta > 0$ this amplification mapping does not depend on the parameter β, an observation from the convergence result in the index-2 case by Jay (2011).

The eigenvalues of the above amplification matrix $\mathbf{T}_\delta$ can analytically be computed as

$$\mu_{2|3} = \frac{\alpha_f - 2\alpha_f\gamma - 2\alpha_m\frac{\delta}{h} + \frac{\delta}{h} + \gamma - 1}{-2((\alpha_f-1)\gamma + (\alpha_m-1)\frac{\delta}{h})} \\ \pm \frac{\sqrt{\alpha_f^2 - 2\alpha_f(\frac{\delta}{h} - \gamma + 1) + 4\alpha_m\frac{\delta}{h} + (\frac{\delta}{h}+\gamma)^2 - 2\frac{\delta}{h} - 2\gamma + 1}}{-2((\alpha_f-1)\gamma + (\alpha_m-1)\frac{\delta}{h})},$$

and in the limit case $\delta \to 0$ resemble the leading error propagation of the index-2 convergence analysis, see Remark 5.16. More importantly, it can be shown that they are—even for all positive values of δ/h—bounded from above by one in absolute value. In the error analysis below we will use an extended version of $\mathbf{T}_\delta$ to cope with the general nonlinear case; also then there is a sufficiently large neighborhood of zero (for values of δ/h) such that the eigenvalues may be bounded, see Example 5.34 below.

In Figure 5.4 the absolute values of the eigenvalues are depicted for the parameter choices from CH(ϱ_∞), HHT and WBZ. Note that for HHT only values $\varrho_\infty \in [0.5, 1]$ are relevant. For $\varrho_\infty \to 0$, $\delta/h = 0$ one eigenvalue would tend towards $-\infty$ if we included that region. The green lines indicate the amplification behavior for $\delta = 0$ and so resemble the plot from Figure 5.3 for the CH(ϱ_∞) method.

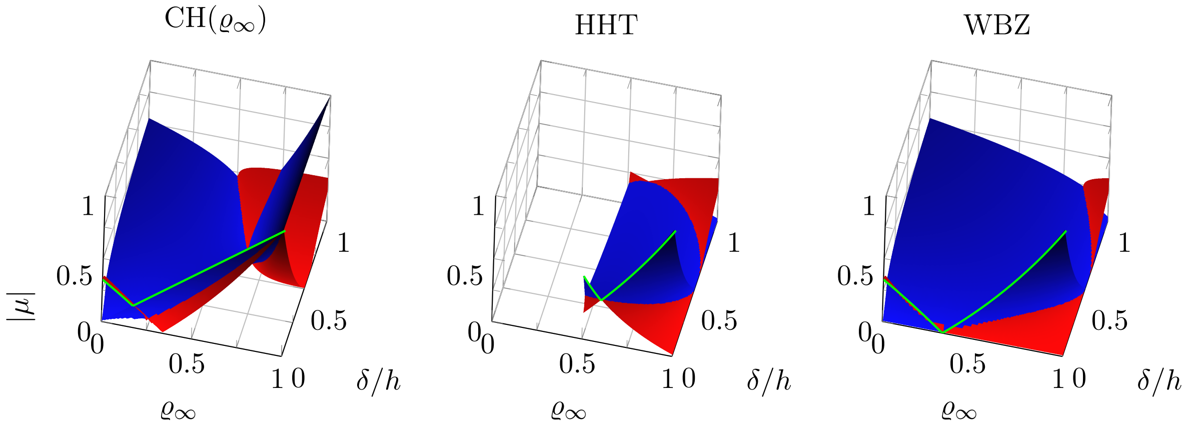


Figure 5.4: Eigenvalues $\mu_{2|3}$ of the linear update mapping from Example 5.27 for CH(ϱ_∞), HHT, and WBZ, green line: behavior for $\delta/h = 0$

Note that, as for the harmonic oscillator in (4.17), rescaling is also possible, i. e., considering the map of $(q, hv, h^2a)^\top$ from one time step to the next one. That way, the representation is closer to classical analysis of Chung and Hulbert (1993) in the case of stiff systems. ♢

Before we proceed with the next, more complex, example we introduce the equivalent error term to (5.8). Motivated by the corresponding singularly perturbed index-1 problem (3.7) and its discrete counterpart in (5.25) we define the global error of the Lagrange multiplier approximations for strongly damped mechanical systems by

$$\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta} := \boldsymbol{\lambda}(t_n) - \boldsymbol{\lambda}_n^\delta , \quad \boldsymbol{\lambda}_n^\delta = \frac{1}{\delta}\mathbf{G}(\boldsymbol{q}_n^\delta)\boldsymbol{v}_n^\delta . \tag{5.27}$$

Equation (5.9) can be extended to the errors $\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}$ in a straightforward manner.

Example 5.28 (Damped Prothero–Robinson equation)
Adding nonlinearity to the solution of the above test problem while keeping the small space dimension one, can be achieved at the cost of having to cope with rheonomic weak constraints. A very similar test problem for first order ODEs has been proposed by Prothero and Robinson (1974) in the analysis of B-convergence for Runge–Kutta methods. An adaptation to second order equations for strongly damped mechanical systems reads (Simeon, 2013)

$$\ddot{q}^\delta(t) = -\frac{1}{\delta}(\dot{q}^\delta(t) - \dot{\varphi}(t)) + \ddot{\varphi}(t) , \quad t \geq 0 , \tag{5.28}$$

with the analytic solution $q^\delta(t) = \varphi(t) + C_1 - \delta C_2\, \mathrm{e}^{-t/\delta}$, where $C_1 := q^\delta(0) - \varphi(0) + \delta(\dot{q}^\delta(0) - \dot{\varphi}(0))$, $C_2 := \dot{q}^\delta(0) - \dot{\varphi}(0)$, and $\varphi\colon \mathbb{R} \to \mathbb{R}$ is an arbitrarily given smooth function. With respect to the general modeling process of Section 3.2, equation (5.28) can generically be obtained from the Rayleigh function $\mathcal{D} := \frac{1}{2}\|\dot{q}^\delta(t) - \dot{\varphi}(t)\|_2^2$, i. e., the *rheonomic* constraint function $g(q,t) := q - \varphi(t)$. For $\varphi \equiv 0$, this is exactly the damped oscillator from Example 5.27. In any case, the slow solution is given by $\boldsymbol{q}(t) = \varphi(t)$ and all errors coincide with the orthogonal error components $\boldsymbol{e}_n^{\mathbf{G}\bullet}$, i. e., the projections and projectors define the DAE solution.

Motivated by Assumption 5.25, for studying one step of (4.2) for (5.28) we assume that the initial values of the SPP are $q^\delta(0) = q_0^\delta = \varphi(0) + C_q h^2$, $\dot{q}^\delta(0) = v_0^\delta = \dot{\varphi}(0) + C_v \delta$ with two constants $C_q, C_v \in \mathbb{R}$. The solution of one step can analytically be computed but is rather cumbersome. Instead, we start by taking a look at the initial error terms to obtain the first distinct difference to the DAE case: For position and velocity level the errors are defined by the chosen initial values. On acceleration-like level and for the original initialization (4.35) we get

$$e_0^{\boldsymbol{a}} = \ddot{\varphi}(\Delta_\alpha h) - a_0^\delta = \ddot{\varphi}(\Delta_\alpha h) - (-\frac{1}{\delta}(v_0^\delta - \dot{\varphi}(0)) + \ddot{\varphi}(0)) = C_v + \mathcal{O}(h) ,$$

and similarly on the level of Lagrange multipliers

$$e_0^{\boldsymbol{\lambda},\delta} = \underbrace{0}_{\equiv \lambda(t)} - (-\frac{1}{\delta}(v_0^\delta - \dot{\varphi}(0))) = C_v .$$

We conclude that these error terms can no longer be estimated by any step size dependent value and will therefore scale the error terms $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}$ and $\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}$ in the analysis below.

The advantage of using the Prothero–Robinson problem as test equation was underlined by Simeon (1998) and Schaub and Simeon (2002) in the context of Runge–Kutta and Rosenbrock methods for stiff mechanical systems: Its very simple structure suffices to describe the methods in the context of stiff force terms while it is also possible to show the sources of possible order reduction. One main result is that *local* (truncation) errors dominate the overall error behavior since global errors are damped out by the algorithms, see also Example 5.35 below. As a result, the explicit error recursion from one time integration step typically suffices for the analysis. The multistep character of Newmark integrators and the fact that acceleration-like variables are, as just seen, no proper error measure necessitate a reformulation as the one already announced in Remark 4.6. Instead of $(e_n^{\boldsymbol{q}}, he_n^{\boldsymbol{v}}, h^2 e_n^{\boldsymbol{a}})^\top$ we phrase the error propagation in terms of $(e_{n+1}^{\boldsymbol{q}}, e_n^{\boldsymbol{q}}, he_n^{\boldsymbol{v}})^\top$. The result is an update

$$\underbrace{\begin{pmatrix} 0 & 0 & * \\ 0 & 1 & 0 \\ * & 0 & * \end{pmatrix}}_{=:\tilde{\mathbf{T}}_1} \underbrace{\begin{pmatrix} e_{n+2}^{\boldsymbol{q}} \\ e_{n+1}^{\boldsymbol{q}} \\ he_{n+1}^{\boldsymbol{v}} \end{pmatrix}}_{=:\tilde{\boldsymbol{E}}_{n+1}} = \underbrace{\begin{pmatrix} * & * & * \\ 1 & 0 & 0 \\ * & * & * \end{pmatrix}}_{=:\tilde{\mathbf{T}}_2} \underbrace{\begin{pmatrix} e_{n+1}^{\boldsymbol{q}} \\ e_n^{\boldsymbol{q}} \\ he_n^{\boldsymbol{v}} \end{pmatrix}}_{=:\tilde{\boldsymbol{E}}_n} + \begin{pmatrix} \eta_n^{\boldsymbol{q}} \\ 0 \\ \eta_n^{\boldsymbol{v}} \end{pmatrix},$$

where the matrix entries marked by an asterisk depend on the four parameters and $z := h/\delta$, cf. Section 4.2.1 and (Erlicher et al. (2002), Kettmann (2009)) for the explicit values. We can obtain the local truncation errors from the corresponding form $\tilde{\boldsymbol{E}}_{n+1} = \tilde{\mathbf{T}}_1^{-1}\tilde{\mathbf{T}}_2\boldsymbol{E}_n + \tilde{\mathbf{T}}_1^{-1}(\eta_n^{\boldsymbol{q}}, 0, \eta_n^{\boldsymbol{v}})^\top$ and find after a series of manipulations, that the third component of the second summand is given by

$$\begin{aligned} \frac{2((\varrho_\infty - 1)\varrho_\infty + 1)}{3((\varrho_\infty - 1)z + 2\varrho_\infty(\varrho_\infty + 1))} \dddot{\varphi}(t_n) \cdot h^3 + \mathcal{O}(\delta h^3) + \mathcal{O}(h^4) &\text{ in case of CH}(\varrho_\infty,) \\ \frac{(3\varrho_\infty - 1)((\varrho_\infty - 1)\varrho_\infty + 1)}{3(\varrho_\infty((\varrho_\infty - 1)^2 z - \varrho_\infty(\varrho_\infty + 3) - 1) + 1)} \dddot{\varphi}(t_n) \cdot h^3 + \mathcal{O}(\delta h^3) + \mathcal{O}(h^4) &\text{ for HHT and} \\ \frac{2((\varrho_\infty - 1)\varrho_\infty + 1)}{3((\varrho_\infty - 1)^2 z - 4\varrho_\infty)} \dddot{\varphi}(t_n) \cdot h^3 + \mathcal{O}(\delta h^3) + \mathcal{O}(h^4) &\text{ for WBZ} \end{aligned}$$

and so does not (necessarily) reduce the order of convergence. For the position components a structurally similar but way more complicated term can be derived.

This is no proof yet but nevertheless a strong indicator that for appropriate initial values no order reduction should occur for strongly damped mechanical systems. ♢

As a first result in the general case we obtain the counterpart of Lemma 5.10 for strongly damped systems.

Lemma 5.29
It holds

$$\begin{aligned} -\delta \cdot \boldsymbol{\lambda}_n^\delta &= \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} + \mathsf{R}(\boldsymbol{q}(t_n))(\dot{\boldsymbol{q}}(t_n), \boldsymbol{e}_n^{\boldsymbol{q}}) + \mathcal{O}(h)(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\|), & (5.29a) \\ \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} &= \delta \boldsymbol{e}_n^{\boldsymbol{\lambda},\delta} + \mathcal{O}(\delta) + \mathcal{O}(h^2) + \mathcal{O}(1)\|\boldsymbol{e}_n^{\boldsymbol{q}}\|. & (5.29b) \end{aligned}$$

Proof. From the above definition of $\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}$ we obtain

$$\begin{aligned} -\delta \cdot \boldsymbol{\lambda}_n^\delta &= \delta(\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta} - \boldsymbol{\lambda}(t_n)) \\ &= -\mathbf{G}(\boldsymbol{q}_n^\delta)\boldsymbol{v}_n^\delta = \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} + \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{v}}\| + \int_0^1 \mathsf{R}(\boldsymbol{q}_n^\delta + \vartheta \boldsymbol{e}_n^{\boldsymbol{q}})(\dot{\boldsymbol{q}}(t_n), \boldsymbol{e}_n^{\boldsymbol{q}})\,\mathrm{d}\vartheta \\ &= \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} + \mathsf{R}(\boldsymbol{q}(t_n))(\dot{\boldsymbol{q}}(t_n), \boldsymbol{e}_n^{\boldsymbol{q}}) + \mathcal{O}(h)(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\|), \end{aligned}$$

which is the first assertion. The second one follows from (5.29a), Assumption 5.26 and $\boldsymbol{\lambda}(t) = \mathcal{O}(1)$. □

An important consequence of Lemma 5.29 is that the error terms $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}$ can completely be expressed in terms of the other error components in all following estimates, in particular it holds

$$\|\boldsymbol{e}_n^{\boldsymbol{v}}\| \leq \mathcal{O}(1)(\|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| + \|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}\|) = \mathcal{O}(1)\|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| + \mathcal{O}(\delta)\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\| + \mathcal{O}(1)\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \mathcal{O}(\delta) + \mathcal{O}(h^2).$$

Hence, the collection of higher order terms in the strongly damped SPP case, as in $\boldsymbol{\eta}_n^{(3)}$ for index-3 DAEs,

$$\boldsymbol{\eta}_n^{(\mathrm{SPP}),\delta} := \|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| + h\|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}\| + h^2\|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\| + h^2\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\|$$

does not comprise the orthogonal velocity error components $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}$.

The difference of estimates (5.29a) for two consecutive time steps implies furthermore

$$\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{v}} - \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} = \delta(\boldsymbol{e}_{n+1}^{\boldsymbol{\lambda},\delta} - \boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}) - \delta(\boldsymbol{\lambda}(t_{n+1}) - \boldsymbol{\lambda}(t_n)) + \mathcal{O}(1)\|\boldsymbol{e}_{n+1}^{\boldsymbol{q}} - \boldsymbol{e}_n^{\boldsymbol{q}}\| + \mathcal{O}(h)(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{e}_{n,n+1}^{\boldsymbol{v}}\|)\,,$$

which by using (5.29b) and Lemma 5.4 leads to

$$\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{v}} - \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} = \tfrac{\delta}{h}(h\boldsymbol{e}_{n+1}^{\boldsymbol{\lambda},\delta} - h\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}) + \mathcal{O}(\delta)\|h\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\| + \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(\mathrm{SPP}),\delta} + \mathcal{O}(\delta h) + \mathcal{O}(h^3)\,. \tag{5.30}$$

The corresponding extension of Lemmas 5.7 and 5.8 in the strongly damped case is given by the following lemma.

Lemma 5.30 (Errors on acceleration level: Strongly damped systems)
The error terms $\boldsymbol{e}_n^{\dot{\boldsymbol{v}}|\boldsymbol{a}}$ on acceleration level and $\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}$ in the Lagrange multipliers relate like

$$\boldsymbol{e}_n^{\dot{\boldsymbol{v}}} + [\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}(t_n))\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta} = \mathcal{O}(1)(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\|)\,, \tag{5.31}$$

$$(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} + \alpha_m\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} + \mathcal{O}(1)\|h\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\| = \mathcal{O}(1)(\boldsymbol{\eta}_{n,n+1}^{(\mathrm{SPP}),\delta} + \|h\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\|) + \mathcal{O}(h^2) + \mathcal{O}(\delta)\,, \tag{5.32}$$

$$(1-\alpha_m)h\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \alpha_m h\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} = -(1-\alpha_f)h\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\delta} - \alpha_f h\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\delta} + \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(\mathrm{SPP}),\delta} + \mathcal{O}(h^3)\,.$$

Proof. Relation (5.31) follows from the equilibrium conditions as in the DAE case taking the artificial Lagrange multipliers into account. We have

$$\begin{aligned}
\boldsymbol{e}_n^{\dot{\boldsymbol{v}}} &= \ddot{\boldsymbol{q}}(t_n) - \dot{\boldsymbol{v}}_n^\delta \\
&= \Big([\mathbf{M}^{-1}\boldsymbol{f}](\boldsymbol{q}(t_n), \dot{\boldsymbol{q}}(t_n)) - [\mathbf{M}^{-1}\boldsymbol{f}](\boldsymbol{q}_n^\delta, \boldsymbol{v}_n^\delta)\Big) - [\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}_n^\delta)\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta} \\
&\quad - \Big([\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}(t_n)) - [\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}_n^\delta)\Big)\boldsymbol{\lambda}(t_n) \\
&= -\boldsymbol{e}_n^{\mathbf{M}^{-1}\mathbf{G}^\top\boldsymbol{\lambda},\delta} + \mathcal{O}(1)(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\|)\,,
\end{aligned}$$

and so (5.31). Note that, since the equilibrium condition (4.2d) of the Newmark method is used for the initialization of the algorithm, i. e., it also holds for $n = -1$, relation (5.31) is also valid for $n = 0$. To attain the other two assertions we employ (5.5) and (2.24) such that in the weighted sum of the $\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}$ terms the influence of the $\boldsymbol{\lambda}$ errors reduces to higher order terms in (5.32) as in the proof of Lemma 5.8. Note that the index of $\boldsymbol{e}_\bullet^{\boldsymbol{\lambda},\delta}$ and $\boldsymbol{e}_\bullet^{\mathbf{G}\boldsymbol{a}}$ in (5.32) is solely taken at the n-th time instance and that we did not need any assumptions concerning the boundedness of $\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\|$. □

In Example 5.28 we have already seen that the initial errors on acceleration-like and Lagrange multiplier level cannot be bounded as in the DAE case which is why in (5.30) and (5.32) we scaled $\boldsymbol{e}_{n,n+1}^{\boldsymbol{\lambda},\delta}$ and $\boldsymbol{e}_{n,n+1}^{\mathbf{G}\boldsymbol{a}}$ by the time step size h. In general the following result holds.

Lemma 5.31 (Initial error terms (III))
The errors in the acceleration and (artificial) Lagrange multiplier solution components in the starting values fulfill

$$\boldsymbol{e}_0^{\boldsymbol{q}} = \mathcal{O}(h^2)\,, \qquad \boldsymbol{e}_0^{\boldsymbol{v}} = \mathcal{O}(\delta)\,, \qquad h\boldsymbol{e}_0^{\mathbf{S}\boldsymbol{\lambda},\delta} = \mathcal{O}(h)\,, \qquad \boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}} + h\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(\delta) + \mathcal{O}(h)\,.$$

Proof. The first two identities are roughly a repetition of Assumption 5.25. The third equation follows from definition (5.27)

$$\boldsymbol{e}_0^{\boldsymbol{\lambda},\delta} = \boldsymbol{\lambda}(t_0) - \tfrac{1}{\delta}\mathbf{G}(\boldsymbol{q}_0^\delta)\boldsymbol{v}_0^\delta = \mathcal{O}(1)\,.$$

This will be used to show the last statement: We employ (5.31) to attain

$$\boldsymbol{e}_0^{\boldsymbol{a}} = \ddot{\boldsymbol{q}}(t_0 + \Delta_\alpha h) - \boldsymbol{a}_0^\delta = \boldsymbol{e}_0^{\dot{\boldsymbol{v}}} + \mathcal{O}(h) = -\boldsymbol{e}_0^{\mathbf{M}^{-1}\mathbf{G}^\top\boldsymbol{\lambda},\delta} + \mathcal{O}(h) = \mathcal{O}(1)\,.$$

For the tangential error terms $\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}}$ equation (2.24) gives the stronger estimate that the term is bounded by a first power of h. □

We are now prepared for the final convergence result:

Theorem 5.32 (Error behavior for strongly damped systems)
Let the parameters of (4.2) satisfy

$$\alpha_m < \alpha_f < \frac{1}{2}\,, \tag{5.33}$$

and presume Assumptions 5.3 and 5.25. Then any (classical) second order Newmark integrator (4.2) fulfills the following error estimates

$$\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| \le C(h^2 + \delta)\,, \qquad \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\| \le C(h\varrho^n + h^2 + \delta)\,,$$

where $\varrho \in [0,1)$ is a constant that depends on the parameters α_m, α_f, γ and the ratio δ/h and n satisfies $t_0 + nh \le t_{\text{end}}$. More precisely, on position and velocity level we have the estimate

$$\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\| \le C(\|\boldsymbol{e}_0^{\boldsymbol{q}}\| + \|\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{v}}\| + \delta + h^2 + h\|\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}}\| + h^2\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}\| + h^2\|\boldsymbol{e}_0^{\boldsymbol{\lambda},\delta}\|)\,.$$

Corollary 5.33 (Convergence for strongly damped systems)
For parameters satisfying (5.33) and initial values fulfilling $\boldsymbol{g}(\boldsymbol{q}_0^\delta) = \mathcal{O}(h^2)$, $\mathbf{G}(\boldsymbol{q}_0^\delta)\boldsymbol{v}_0^\delta = \mathcal{O}(\delta)$ there exist initial values $(\boldsymbol{q}(t_0), \dot{\boldsymbol{q}}(t_0))^\top$ with $\boldsymbol{g}(\boldsymbol{q}(t_0)) = \mathbf{G}(\boldsymbol{q}(t_0))\dot{\boldsymbol{q}}(t_0) = \mathbf{0}$ and with differences $\boldsymbol{q}(t_0) - \boldsymbol{q}_0^\delta$, $\dot{\boldsymbol{q}}(t_0) - \boldsymbol{v}_0^\delta$ in the $\mathbf{M}(\boldsymbol{q}(t_0))$-orthogonal complement of the tangential space of $\mathfrak{M}^{\mathrm{s}}$ in $\boldsymbol{q}(t_0)$ such that for $\delta < C_0 h$ and $C_0 > 0$ sufficiently small, the numerical approximations of the Newmark integrator (4.2) satisfy

$$\|\boldsymbol{q}(t_n) - \boldsymbol{q}_n^\delta\| + \|\dot{\boldsymbol{q}}(t_n) - \boldsymbol{v}_n^\delta\| \le C(\delta + h^2)\,,$$

whenever $t_n = t_0 + nh \le t_{\text{end}}$. The constant $C > 0$ is independent of h, δ and n and $\boldsymbol{q}(t)$, $t \in [t_0, t_{\text{end}}]$, denotes the solution of the DAE system (2.13) with initial values $\boldsymbol{q}(t_0)$, $\dot{\boldsymbol{q}}(t_0)$.

Proof of Theorem 5.32. For the application of Lemma 5.1 we collect the results from Corollary 5.5, in particular (5.6a) and (5.6c) to obtain with (5.29b)

$$\begin{pmatrix} \boldsymbol{e}_{n+1}^{\boldsymbol{q}} \\ \boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{v}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_n^{\boldsymbol{q}} \\ \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}} \end{pmatrix} + \mathcal{O}(h)\left(\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| + \tfrac{\delta}{h}\|h\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\| + \|\boldsymbol{e}_{n,n+1}^{\mathbf{P}\boldsymbol{a}}\| + \|h\boldsymbol{e}_{n,n+1}^{\mathbf{G}\boldsymbol{a}}\|\right) \\ + \mathcal{O}(h^3) + \mathcal{O}(h\delta)\,, \tag{5.34}$$

which will provide the first set of inequalities (5.1a). For the vector-valued estimate (5.1b) we collect our findings from Lemma 5.30 and combine the estimates from (5.6d) and (5.30) to obtain

$$\overbrace{\begin{pmatrix} (1-\alpha_m)\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\delta}{h}\mathbf{S}_n^{-1} & -\gamma\mathbf{I} \\ \mathbf{0} & (1-\alpha_f)\mathbf{I} & (1-\alpha_m)\mathbf{I} \end{pmatrix}}^{:=\mathbf{T}_{1,\delta}(\delta/h\mathbf{S}_n^{-1})} \begin{pmatrix} \boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} \\ h\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\delta} \\ h\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} \end{pmatrix} = \overbrace{\begin{pmatrix} -\alpha_m\mathbf{I} & \mathcal{O}(1) & \mathcal{O}(1) \\ \mathbf{0} & \frac{\delta}{h}\mathbf{S}_n^{-1} & (1-\gamma)\mathbf{I} \\ \mathbf{0} & -\alpha_f\mathbf{I} & -\alpha_m\mathbf{I} \end{pmatrix}}^{:=\mathbf{T}_{2,\delta}(\delta/h\mathbf{S}_n^{-1}} \begin{pmatrix} \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} \\ h\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\delta} \\ h\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} \end{pmatrix} \\ + \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(\mathrm{SPP}),\delta} + \mathcal{O}(h^2) + \mathcal{O}(\delta)\,. \tag{5.35}$$

The matrix $\mathbf{S}_n := \mathbf{S}(\boldsymbol{q}(t_n))$ has already been defined in (2.20) and is symmetric and positive definite for all $n \geq 0$ which is why the error terms $\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}$ and $\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\delta}$ can be treated almost equivalently. Note that we use the same index on both sides of the equation which is justified because of

$$\frac{\delta}{h}\mathbf{S}_{n+1}^{-1} h \boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\delta} = \delta(\mathbf{S}_{n+1}^{-1} - \mathbf{S}_n^{-1})\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\delta} + \delta \mathbf{S}_n^{-1} \boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\delta} = \frac{\delta}{h}\mathbf{S}_n^{-1} h \boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\delta} + \mathcal{O}(1)\boldsymbol{\eta}_{n+1}^{(\mathrm{SPP}),\delta} .$$

The occurrence of time step dependent entries in the error amplification formulae is the next fundamental difference to the DAE case where only constant matrices govern the leading error terms. We also notice that the lower two-by-two block of the leading terms in (5.35) resembles the mapping $\mathbf{T}_\delta$ in the linear case from Example 5.27. Its eigenvalues have already been discussed and shown to be smaller than one in absolute value (for $n_{\boldsymbol{q}} = 1$). The first line in (5.35) is also basically known from the DAE case for both, index-2 and index-3, but differs in the additional $(1,2)$ and $(1,3)$ block entries which emerge from the rescaling of the error terms. A decoupling of the lower two-by-two block can be achieved if we define for each n the orthogonal matrix $\mathbf{U}_n$ that diagonalizes $\mathbf{S}_n^{-1}$ like

$$\mathbf{U}_n \mathbf{S}_n^{-1} \mathbf{U}_n^\top =: \boldsymbol{\Lambda}_n^{-1} = \mathrm{diag}((\mu_i(\mathbf{S}_n^{-1}))_{i=1,\ldots,n_{\boldsymbol{\lambda}}}) , \tag{5.36}$$

where $\mu_i(\mathbf{S}_n^{-1})$, $i = 1, \ldots, n_{\boldsymbol{\lambda}}$, denote the positive eigenvalues of $\mathbf{S}_n^{-1}$. (Due to compactness of $[t_0, t_{\mathrm{end}}]$ and Assumption 5.26 there is a compact neighborhood of the solution $\boldsymbol{q}(t)$ such that these eigenvalues can be bounded from above and below.) So, if from now on instead of $\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\delta}$ and $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}$ we consider $\boldsymbol{e}_n^{\mathbf{U}^\top\mathbf{S}\boldsymbol{\lambda},\delta}$ and $\boldsymbol{e}_n^{\mathbf{U}^\top\mathbf{G}\boldsymbol{a}}$ the equations can (almost) be treated as if they were defined in $\mathbb{R}^3$ rather than $\mathbb{R}^{3n_{\boldsymbol{\lambda}}}$. The only exceptions are the block entries on the right hand side but the structure of both block matrices implies that

$$\mathbf{T}(\tfrac{\delta}{h}\mathbf{S}_n^{-1}) := \mathbf{T}_{1,\delta}^{-1}(\tfrac{\delta}{h}\mathbf{S}_n^{-1}) \cdot \mathbf{T}_{2,\delta}(\tfrac{\delta}{h}\mathbf{S}_n^{-1}) = \begin{pmatrix} \frac{-\alpha_m}{1-\alpha_m}\mathbf{I} & \mathcal{O}(1) & \mathcal{O}(1) \\ \mathbf{0} & \multicolumn{2}{c}{\multirow{2}{*}{$\tilde{\mathbf{T}}(\tfrac{\delta}{h}\mathbf{S}_n^{-1})$}} \\ \mathbf{0} & & \end{pmatrix} ,$$

where $\tilde{\mathbf{T}}(\frac{\delta}{h}\mathbf{S}_n^{-1})$ is the according result for only the lower two-by-two block which can be decoupled using the above procedure.

In Lemma 5.1 we used a constant matrix $\mathbf{T}$ for simplicity reasons only; an extension to variable matrices is easily done. The crucial requirement, however, is that the involved vector norms (and the corresponding matrix norm) remains the same which is not encompassed by the above result where the existence of a suitable norm is only guaranteed for fixed matrices. So, our next goal is to construct a norm $\|\bullet\|_*$ which corresponds to a vector norm and fulfills

$$\forall n : \|\mathbf{T}(\tfrac{\delta}{h}\mathbf{S}_n^{-1})\|_* < 1 . \tag{5.37}$$

The eigenvalues of $\tilde{\mathbf{T}}(\frac{\delta}{h}\mathbf{S}_n^{-1})$ can be bounded by one if $\frac{\delta}{h}$ is sufficiently small, cf. Example 5.27 and also Example 5.34 below. It is a well-known result, cf. (Hairer et al., 1993, Lemma III.4.4), that for each matrix $\tilde{\mathbf{T}}$ with spectral radius $\varrho(\tilde{\mathbf{T}})$ and each positive number $\kappa > 0$ there exists a matrix norm $\|\bullet\|_\circ$ which corresponds to a vector norm such that

$$\varrho(\tilde{\mathbf{T}}) \leq \|\tilde{\mathbf{T}}\|_\circ + \kappa .$$

The proof uses the transformation to Jordan canonical form which for the matrices $\tilde{\mathbf{T}}(\frac{\delta}{h}\mathbf{S}_n^{-1})$ is a transformation to diagonal form

$$\tilde{\mathbf{T}} =: \tilde{\mathbf{C}} \ \mathrm{diag}(\lambda_i(\tilde{\mathbf{T}})) \, \tilde{\mathbf{C}}^{-1} .$$

We define the vector norm $\|\bullet\|_\circ := \|\tilde{\mathbf{C}}^{-1}\bullet\|_\infty$ which corresponds to the matrix norm $\|\bullet\|_\circ := \|\tilde{\mathbf{C}}^{-1} \bullet \tilde{\mathbf{C}}\|_\infty$ such that, e. g. for $\frac{\delta}{h} = 0$, we get the existence of such a norm with $\|\tilde{\mathbf{T}}(\mathbf{0})\|_\circ < 1$. We now use that $\mathbf{T}(\frac{\delta}{h}\mathbf{S}_n^{-1})$ depends *smoothly* on $\frac{\delta}{h}$. Remembering Assumption 5.3 that the penalty parameter δ is smaller than $C_0 h$ for a (possibly very small) fixed number $C_0 > 0$ we can now use the matrix norm for $\frac{\delta}{h} = 0$ and its continuity to obtain $\|\tilde{\mathbf{T}}(\frac{\delta}{h}\mathbf{S}_n^{-1})\|_\circ < 1$ for all $n \geq 0$.

To extend this norm to $\mathbb{R}^{n_{\boldsymbol{q}}+2n_{\boldsymbol{\lambda}}}$ we use that (5.33) implies that $|\frac{-\alpha_m}{1-\alpha_m}| < 1$. If we define the matrix $\mathbf{C} := \operatorname{blkdiag}(\varkappa\mathbf{I}_{n_{\boldsymbol{q}}}, \tilde{\mathbf{C}})$ with a sufficiently small constant $0 < \varkappa < 1$, then the matrix norm $\|\bullet\|_* := \|\mathbf{C}^{-1} \bullet \mathbf{C}\|_\infty$ on $\mathbb{R}^{n_{\boldsymbol{q}}+2n_{\boldsymbol{\lambda}}}$ fulfills (5.37) whenever the conditions on the parameters are valid because

$$\mathbf{C}\mathbf{T}(\tfrac{\delta}{h}\mathbf{S}_n^{-1})\mathbf{C}^{-1} = \begin{pmatrix} \frac{-\alpha_m}{1-\alpha_m}\mathbf{I}_{n_{\boldsymbol{q}}} & \mathcal{O}(\varkappa) \quad \mathcal{O}(\varkappa) \\ \mathbf{0} & \multirow{2}{*}{$\tilde{\mathbf{T}}(\tfrac{\delta}{h}\mathbf{S}_n^{-1})$} \\ \mathbf{0} & \end{pmatrix}.$$

In conclusion, as a result of Lemma 5.1 and (5.35) we obtain with $\boldsymbol{E}_n^{\boldsymbol{y}} := (\boldsymbol{e}_n^{\boldsymbol{q}}, \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}})^\top$ and $\boldsymbol{E}_n^{\boldsymbol{z}} := (\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}, h\boldsymbol{e}_n^{\mathbf{U}^\top\mathbf{S}\boldsymbol{\lambda},\delta}, h\boldsymbol{e}_n^{\mathbf{U}^\top\mathbf{G}\boldsymbol{a}})^\top$ the error bounds

$$\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| \leq C(h^2 + \delta)\,, \qquad \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}\| \leq C(h\varrho^n + h^2 + \delta)$$

with a suitable constant $\varrho \in [0, 1)$. Finally, we can use (5.29b) one last time to use the convergence result for $\boldsymbol{q}$ and $\boldsymbol{\lambda}$ such that the upper bound

$$\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} = \delta \underbrace{\boldsymbol{e}_n^{\boldsymbol{\lambda},\delta}}_{=\mathcal{O}(1)} + \mathcal{O}(\delta) + \mathcal{O}(h^2) + \mathcal{O}(1) \underbrace{\|\boldsymbol{e}_n^{\boldsymbol{q}}\|}_{=\mathcal{O}(h^2)+\mathcal{O}(\delta)} = \mathcal{O}(\delta) + \mathcal{O}(h^2)\,,$$

follows and so Theorem 5.32. □

Note that, as in the index-2 case, no conditions on the parameter β are involved in the convergence result for strongly damped systems. For just contractivity of the above mapping one would also have to imply that $\gamma \geq \frac{1}{2}$ but with (5.33) and the second order condition (4.3) this is automatically fulfilled. Although the proof does not explicitly involve zero stability of the underlying ODE method as it relies on the onestep representation condition (4.18) is still ensured. Note also that due to the $\mathcal{O}(1)$-elements in the propagation matrix $\mathbf{T}$ the additional separation of error terms as in the DAE case is not possible any longer.

Example 5.34 (Norm for the CH(ϱ_∞) algorithm)
To gain a better understanding of the above requirement that the constant C_0 in Assumption 5.3 needs to be sufficiently small (depending on the parameters and the solution of the slow system) to guarantee contractivity in one unified norm, we consider again the parameter choice of Chung and Hulbert (1993). For simplicity only we take $n_{\boldsymbol{q}} = 1$ and the above matrix $\tilde{\mathbf{T}}$ takes the form

$$\tilde{\mathbf{T}}(\tfrac{\delta}{h}s^{-1}) := \begin{pmatrix} \frac{2\frac{\delta}{h}s^{-1}(\varrho_\infty-2)(\varrho_\infty+1)-(\varrho_\infty-3)\varrho_\infty}{\varrho_\infty+2\frac{\delta}{h}s^{-1}(\varrho_\infty-2)(\varrho_\infty+1)-3} & \frac{\varrho_\infty^2-1}{\varrho_\infty+2\frac{\delta}{h}s^{-1}(\varrho_\infty-2)(\varrho_\infty+1)-3} \\ \frac{2\frac{\delta}{h}s^{-1}(\varrho_\infty+1)^2}{\varrho_\infty+2\frac{\delta}{h}s^{-1}(\varrho_\infty-2)(\varrho_\infty+1)-3} & \frac{\varrho_\infty+6\frac{\delta}{h}s^{-1}(\varrho_\infty+1)+5}{\varrho_\infty+2\frac{\delta}{h}s^{-1}(\varrho_\infty-2)(\varrho_\infty+1)-3} + 2 \end{pmatrix}. \tag{5.38}$$

If we take the limit case $\delta = 0$ then $\tilde{\mathbf{T}}$ reduces to the amplification matrix from the proof in the index-2 case given in Remark 5.16. For the construction of a suitable vector norm we diagonalize this matrix using

$$\tilde{\mathbf{C}} := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{such that} \quad \tilde{\mathbf{C}}^{-1} \cdot \tilde{\mathbf{T}}(0) \cdot \tilde{\mathbf{C}} = \operatorname{diag}(\mu_1, \mu_2)$$

with the eigenvalues $-1 \leq \mu_i \leq 1$, $i = 1, 2$, of $\tilde{\mathbf{T}}(0)$. As the infinity norm of $\operatorname{diag}(\mu_1, \mu_2)$ is bounded by one, we use the matrix norm

$$\|\bullet\|_\circ := \|\tilde{\mathbf{C}}^{-1} \cdot \bullet \cdot \tilde{\mathbf{C}}\|_\infty .$$

In Figure 5.5 we present a numerical evaluation of the $\|\bullet\|_\circ$ norm for varying values of $\varrho_\infty \in [0, 1]$

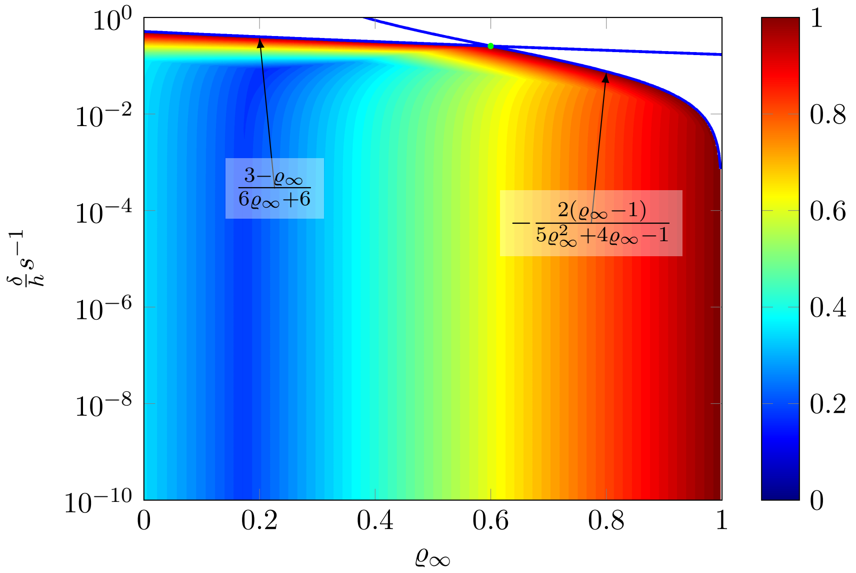


Figure 5.5: Norm of lower two-by-two block-matrix in (5.38), only values smaller one displayed

and $\frac{\delta}{h} s^{-1}$. For this choice one can show that the above norm is strictly bounded by one whenever

$$\frac{\delta}{h} s^{-1} < \begin{cases} \dfrac{3 - \varrho_\infty}{6\varrho_\infty + 6} & \text{for } \varrho_\infty \in [0, 0.6]\,, \\ -\dfrac{2(\varrho_\infty - 1)}{5\varrho_\infty^2 + 4\varrho_\infty - 1} & \text{for } \varrho_\infty \in [0.6, 1]\,. \end{cases}$$

Note also that this bound is rather pessimistic as we display just one particular choice of a norm and that the overall error propagation additionally involves the first $n_{\boldsymbol{q}}$ rows in (5.35). $\diamondsuit$

5.3.2 Stiff mechanical systems

In the above section we have outlined the three major differences from the viewpoint of numerical analysis between the convergence analysis in the case of constrained systems on the one and SPPs on the other hand: First, the large initial errors on level of accelerations and/or Lagrange multipliers necessitate that the according error terms are appropriately scaled, second, that the introduction of penalty force terms implies that constraint violations are no longer negligible and so additional coupling terms are present which, third, result in a time step dependent overall error amplification. All three points appear as well for stiff mechanical systems. A first convergence result for Newmark integration methods in the present form and stiff mechanical systems was published in (Köbis and Arnold, 2014). In this section we follow and extend the representation in (Köbis and Arnold, 2016). One step of the algorithm in case of the numerical solution of stiff mechanical systems (3.15) or their index-1 formulation (3.16) reads

$$\begin{array}{rrcl} & \boldsymbol{q}^\varepsilon_{n+1} &=& \boldsymbol{q}^\varepsilon_n + h\boldsymbol{v}^\varepsilon_n + h^2(\tfrac{1}{2}-\beta)\boldsymbol{a}^\varepsilon_n + h^2\beta\boldsymbol{a}^\varepsilon_{n+1}\,, \\ & \boldsymbol{v}^\varepsilon_{n+1} &=& \boldsymbol{v}^\varepsilon_n + h(1-\gamma)\boldsymbol{a}^\varepsilon_n + h\gamma\boldsymbol{a}^\varepsilon_{n+1}\,, \\ & (1-\alpha_m)\boldsymbol{a}^\varepsilon_{n+1} + \alpha_m\boldsymbol{a}^\varepsilon_n &=& (1-\alpha_f)\dot{\boldsymbol{v}}^\varepsilon_{n+1} + \alpha_f\dot{\boldsymbol{v}}^\varepsilon_n\,, \\ \text{a) } \{ & \mathbf{M}(\boldsymbol{q}^\varepsilon_{n+1})\dot{\boldsymbol{v}}^\varepsilon_{n+1} &=& \boldsymbol{f}(\boldsymbol{q}^\varepsilon_{n+1},\boldsymbol{v}^\varepsilon_{n+1}) - \frac{1}{\varepsilon^2}\mathbf{G}^\top(\boldsymbol{q}^\varepsilon_{n+1})\boldsymbol{g}(\boldsymbol{q}^\varepsilon_{n+1})\,, \\ \text{or b) } \left\{ \begin{array}{l} \\ \\ \end{array}\right. & \begin{array}{r} \mathbf{M}(\boldsymbol{q}^\varepsilon_{n+1})\dot{\boldsymbol{v}}^\varepsilon_{n+1} \\ \varepsilon^2\boldsymbol{\lambda}^\varepsilon_{n+1} \end{array} & \begin{array}{c} = \\ = \end{array} & \begin{array}{l} \boldsymbol{f}(\boldsymbol{q}^\varepsilon_{n+1},\boldsymbol{v}^\varepsilon_{n+1}) - \mathbf{G}^\top(\boldsymbol{q}^\varepsilon_{n+1})\boldsymbol{\lambda}^\varepsilon_{n+1}\,, \\ \boldsymbol{g}(\boldsymbol{q}^\varepsilon_{n+1})\,. \end{array} \end{array} \tag{5.39}$$

In a first step we will again define a new set of (for most practical applications artificial) error terms, i. e.,

$$\boldsymbol{e}^{\boldsymbol{\lambda},\varepsilon}_n := \boldsymbol{\lambda}(t_n) - \boldsymbol{\lambda}^\varepsilon_n\,, \quad \boldsymbol{\lambda}^\varepsilon_n = \frac{1}{\varepsilon^2}\boldsymbol{g}(\boldsymbol{q}^\varepsilon_n)\,.$$

Again, we will start with a one-dimensional example before the general theory is established.

Example 5.35 (Second order Prothero–Robinson problem: The stiff case)
We consider the scalar test equation

$$\ddot{q}^\varepsilon(t) = -\frac{1}{\varepsilon^2}(q^\varepsilon(t) - \varphi(t)) + \ddot{\varphi}(t)\,, \quad t \geq 0\,, \tag{5.40}$$

The (smooth) function $\varphi\colon \mathbb{R} \to \mathbb{R}$ can be chosen freely and is at the same time the smooth component of the analytic solution which is given by

$$q^\varepsilon(t) = \varphi(t) + \varepsilon\Delta_v \sin((t-t_0)\varepsilon^{-1}) + \Delta_q \cos((t-t_0)\varepsilon^{-1})$$

with $\Delta_q = q^\varepsilon(0) - \varphi(0)$, $\Delta_v = \dot{q}^\varepsilon(0) - \dot{\varphi}(0)$. The problem has first been considered by van der Houwen and Sommeijer (1987) in the context of Runge–Kutta–Nystöm methods as an extension of the 'classical' Prothero–Robinson test problem (Prothero and Robinson, 1974). It has also been used for the analysis of Runge–Kutta and Rosenbrock-type methods, see (Scholz, 1989, Simeon, 1998, Becker et al., 2014). As its strongly damped counterpart in Example 5.28 the problem is rheonomic and so actually not in the scope of the current investigations but can be obtained by almost exactly the same procedure: Define an additional penalty potential term $\mathcal{U} := \frac{1}{2\varepsilon^2}(\varphi(t) - q^\varepsilon(t))^2$ as in (3.14) and apply Lagrange's formalism. To study the error behavior we consider with Assumption 5.25 that the initial deviations on position and velocity level are in $\mathcal{O}(\varepsilon^2)$, in particular $q^\varepsilon(0) = \varphi(0) + C_q\varepsilon^2$, $C_q \in \mathbb{R}$. With this choice and (4.35) the initial error terms again cannot be bounded by the time step size since

$$e^{\boldsymbol{a}}_0 = \ddot{\varphi}(\Delta_\alpha h) - a^\varepsilon_0 = \ddot{\varphi}(\Delta_\alpha h) - (-\tfrac{1}{\varepsilon^2}C_q\varepsilon^2 + \ddot{\varphi}(0)) = C_q + \mathcal{O}(h)\,, \quad e^{\boldsymbol{\lambda},\varepsilon}_0 = \underbrace{\lambda(0)}_{=0} - \tfrac{C_q\varepsilon^2}{\varepsilon^2} = -C_q\,.$$

Additionally, we can observe that also the order reduction known from the index-3 case carries over: Motivated by Example 5.19 we take a polynomial function $\varphi(t) = t^2$ and perform one step with the above initial values in exact arithmetics. As a result we obtain

$$e^{\boldsymbol{q}}_1 = \varepsilon^2 C_q \cdot \frac{(2(\alpha_m - 1)\frac{\varepsilon^2}{h^2} - (\alpha_m - 2\alpha_f\beta + 2\beta - 1))}{2((\alpha_f - 1)\beta + (\alpha_m - 1)\frac{\varepsilon^2}{h^2})}\,,$$

$$e^{\boldsymbol{v}}_1 = hC_q \cdot \frac{((1-\alpha_f)(2\beta - \gamma) + 2(1-\alpha_m)\frac{\varepsilon^2}{h^2})}{2((\alpha_f - 1)\beta + (\alpha_m - 1)\frac{\varepsilon^2}{h^2})}\,,$$

i. e., a drop of the order to one in the velocity components (at least in $\boldsymbol{e}_1^{\mathbf{G}\boldsymbol{v}}$).

This observation raises the question whether the order reduction is a generic problem or again only a transient phenomenon. A similar procedure as in Example 5.28 for again arbitrary smooth function φ can be used to obtain the update formulae

$$\underbrace{\begin{pmatrix} e_2^{\boldsymbol{q}} \\ e_1^{\boldsymbol{q}} \\ he_1^{\boldsymbol{v}} \end{pmatrix}}_{=:\hat{\boldsymbol{E}}_{n+1}} = \underbrace{\begin{pmatrix} * & * & * \\ 1 & 0 & 0 \\ * & * & * \end{pmatrix}}_{=:\hat{\mathbf{T}}} \underbrace{\begin{pmatrix} e_1^{\boldsymbol{q}} \\ e_0^{\boldsymbol{q}} \\ he_0^{\boldsymbol{v}} \end{pmatrix}}_{=:\tilde{\boldsymbol{E}}_n} + \underbrace{\begin{pmatrix} \hat{\eta}_0^{\boldsymbol{q}} \\ 0 \\ \hat{\eta}_0^{\boldsymbol{v}} \end{pmatrix}}_{=:\hat{\boldsymbol{\eta}}_0},$$

with a matrix $\hat{\mathbf{T}}$ that has the same eigenvalues as the amplification matrix of the harmonic oscillator example from (4.16). The global errors from previous time steps are therefore rapidly damped out and their main origin stems from the local error terms $\hat{\eta}_0^{\boldsymbol{q}|\boldsymbol{v}}$ as it is one main result of the analysis in the case of Runge–Kutta methods by Simeon (1998). Their explicit structure is rather complicated. For CH(0.75) it holds

$$\hat{\boldsymbol{\eta}}_0 = \begin{pmatrix} (192 + 735\frac{\varepsilon^2}{h^2}) \cdot \left(-\frac{182}{3}h^3\varepsilon^2\dddot{\varphi}(t_n)\right) \\ 0 \\ -\frac{13}{63}h^3\dddot{\varphi}(t_n) \end{pmatrix} + \mathcal{O}(h^4),$$

such that—once a transient phase has been surpassed (and only based on this example)—no further drop in order should be expected and the numerically observable order of the global error after sufficiently many steps should be two apart from the 'modeling error' of the singularly perturbed formulation which remains in the magnitude of the penalty parameter ε. ♢

Within the analysis of stiff mechanical systems the orthogonal velocity error terms $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}$ cannot be eliminated as for strongly damped systems. So, the condensed higher order error terms take the form

$$\boldsymbol{\eta}_n^{(\mathrm{SPP}),\varepsilon} := \|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| + h\left(\|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}\|\right) + h^2\left(\|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\| + \|\boldsymbol{e}_n^{\boldsymbol{\lambda},\varepsilon}\|\right).$$

The coupling of position constraints to the dynamic equations in the stiff case implies that the equivalent to (5.30) for stiff mechanical systems does not involve the velocity components. Instead, one gets the following interrelation of errors on multiplier level and $\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{q}}$.

Lemma 5.36
It holds

$$\mathbf{G}(\boldsymbol{q}(t_n))\,\boldsymbol{\Delta}_h\,\boldsymbol{e}_n^{\boldsymbol{q}} = \frac{\varepsilon^2}{h^2}\left(h\boldsymbol{e}_{n+1}^{\boldsymbol{\lambda},\varepsilon} - h\boldsymbol{e}_n^{\boldsymbol{\lambda},\varepsilon}\right) + \mathcal{O}(1)\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\| + \mathcal{O}(\varepsilon^2).$$

Proof. Following the proof of Lemma 4 in (Arnold et al., 2015a) in a first step we observe

$$\begin{aligned}\mathbf{G}(\boldsymbol{q}(t_n))\boldsymbol{e}_n^{\boldsymbol{q}} &= \int_0^1 \left(\mathbf{G}(\boldsymbol{q}(t_n)) - \mathbf{G}(\boldsymbol{q}(t_n) - \vartheta\boldsymbol{e}_n^{\boldsymbol{q}})\right)\boldsymbol{e}_n^{\boldsymbol{q}}\,\mathrm{d}\vartheta + \int_0^1 \mathbf{G}(\boldsymbol{q}(t_n) - \vartheta\boldsymbol{e}_n^{\boldsymbol{q}})\boldsymbol{e}_n^{\boldsymbol{q}}\,\mathrm{d}\vartheta \\ &= \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \underbrace{\boldsymbol{g}(\boldsymbol{q}(t_n))}_{=\mathbf{0}} - \boldsymbol{g}(\boldsymbol{q}_n^{\varepsilon}) = \mathcal{O}(h)\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \varepsilon^2\boldsymbol{e}_n^{\boldsymbol{\lambda},\varepsilon} - \varepsilon^2\boldsymbol{\lambda}(t_n).\end{aligned}$$

Subtracting this estimate for two successive time instances and scaling by $\frac{1}{h}$ gives

$$\mathbf{G}(\boldsymbol{q}(t_n))\frac{\boldsymbol{e}_{n+1}^{\boldsymbol{q}} - \boldsymbol{e}_n^{\boldsymbol{q}}}{h} = -\varepsilon^2\underbrace{\frac{\boldsymbol{\lambda}(t_{n+1}) - \boldsymbol{\lambda}(t_n)}{h}}_{=\dot{\boldsymbol{\lambda}}(t_n)+\mathcal{O}(h)=\mathcal{O}(1)} + \frac{\varepsilon^2}{h^2}\left(h\boldsymbol{e}_{n+1}^{\boldsymbol{\lambda},\varepsilon} - h\boldsymbol{e}_n^{\boldsymbol{\lambda},\varepsilon}\right) + \mathcal{O}(1)\|\boldsymbol{e}_{n,n+1}^{\boldsymbol{q}}\|,$$

which completes the proof. □

For the acceleration level variables we obtain a very similar result as in Lemma 5.30. With the new condensed error terms these estimates read like given in the following lemma.

Lemma 5.37
The error components on acceleration level in tangential and orthogonal direction to the constraint manifold fulfill

$$\begin{aligned} \boldsymbol{e}_n^{\dot{\boldsymbol{v}}} &= -[\mathbf{M}^{-1}\mathbf{G}^\top](\boldsymbol{q}(t_n))\boldsymbol{e}_n^{\boldsymbol{\lambda},\varepsilon} + \mathcal{O}(1)\left(\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\boldsymbol{v}}\|\right) , \\ \boldsymbol{e}_n^{\mathbf{P}\dot{\boldsymbol{v}}} &= \mathcal{O}(1)\boldsymbol{\eta}_n^{(\mathrm{SPP}),\varepsilon} , \\ \boldsymbol{e}_n^{\mathbf{G}\dot{\boldsymbol{v}}} &= -\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} + \mathcal{O}(1)\boldsymbol{\eta}_n^{(\mathrm{SPP}),\varepsilon} . \end{aligned}$$

Proof. As in the proof of (5.31) the results follow from subtracting the equilibrium conditions of the algorithm for the constrained and the singularly perturbed system taking into account the definition of the errors on Lagrange multiplier level and the smoothness of the involved values $\mathbf{M}^{-1}$, $\boldsymbol{f}$ and $\mathbf{G}$. □

Corollary 5.38 (Errors on acceleration level: Stiff mechanical systems)
The stiff equivalent of Lemma 5.30 reads

$$(1-\alpha_m)\boldsymbol{e}_{n+1}^{\mathbf{P}\boldsymbol{a}} + \alpha_m \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} = \mathcal{O}(1)\left(\boldsymbol{\eta}_{n,n+1}^{(\mathrm{SPP}),\varepsilon} + h\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} + h\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\right) + \mathcal{O}(h^2) , \tag{5.41a}$$

$$(1-\alpha_m)h\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} + \alpha_m h\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} = -(1-\alpha_f)h\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} - \alpha_f h\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} + \mathcal{O}(h)\boldsymbol{\eta}_{n,n+1}^{(\mathrm{SPP}),\varepsilon} + \mathcal{O}(h^3) . \tag{5.41b}$$

Proof. Combine (5.5) and Lemma 5.37. □

Lemma 5.39 (Initial error terms (IV))
For Newmark integration methods (4.2) applied to stiff mechanical systems the following estimates for the initial error terms hold

$$\boldsymbol{e}_0^{\boldsymbol{q}} + \boldsymbol{e}_0^{\mathbf{P}\boldsymbol{v}} = \mathcal{O}(\varepsilon^2) , \quad \boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}} = \mathcal{O}(\varepsilon^2) , \quad h\boldsymbol{e}_0^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} = \mathcal{O}(h) , \quad \boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}} + h\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}} = \mathcal{O}(\varepsilon^2) + \mathcal{O}(h) .$$

Proof. The result follows from the same reasoning as the examination of initial error terms in Example 5.35: For the initial errors on Lagrange multiplier level only a $\mathcal{O}(1)$-bound may be obtained presuming only Assumption 5.25. For the errors on acceleration level this fact and Lemma 5.37 give the result as in the proof of Lemma 5.31. □

Theorem 5.40 (Error behavior for stiff mechanical systems)
Let Assumptions 5.3 and 5.25 and the second order condition (4.3) be fulfilled. A stable Newmark integration scheme in the sense of Definition 4.16 for the equations of motion of the stiff mechanical systems (3.15) fulfills the error estimates

$$\|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| \le C(h^2+\varepsilon^2) , \quad \|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}\| + h\|\boldsymbol{e}_n^{\boldsymbol{\lambda},\varepsilon}\| \le C(h\varrho^n + h^2 + \varepsilon^2) ,$$

where $\varrho \in [0,1)$ is a constant that depends on the parameters and the ratio $\frac{\varepsilon}{h}$. More precisely, on position and velocity level we have the estimates

$$\begin{aligned} \|\boldsymbol{e}_n^{\boldsymbol{q}}\| + \|\boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}}\| &\le C(\overbrace{\|\boldsymbol{e}_0^{\boldsymbol{q}}\| + \|\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{v}}\| + h\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}}\| + h\|\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}}\| + h^2\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}\| + h^2\|\boldsymbol{e}_0^{\boldsymbol{\lambda},\varepsilon}\|}^{=\boldsymbol{\eta}_0^{(\mathrm{SPP}),\varepsilon}} + h^2 + \varepsilon^2) , \\ \|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}\| &\le C\left(\varrho^n(\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{v}}\| + \|\boldsymbol{e}_0^{\mathbf{P}\boldsymbol{a}}\| + h\|\boldsymbol{e}_0^{\mathbf{G}\boldsymbol{a}}\| + h\|\boldsymbol{e}_0^{\boldsymbol{\lambda},\varepsilon}\|) + \boldsymbol{\eta}_0^{(\mathrm{SPP}),\varepsilon} + h^2 + \varepsilon^2\right) . \end{aligned}$$

Corollary 5.41 (see Theorem 1 in (Köbis and Arnold, 2014, Köbis and Arnold, 2016))
For numerical damping parameter $\varrho_\infty \in [0,1)$ and initial values satisfying $\boldsymbol{g}(\boldsymbol{q}_0^\varepsilon) = \mathcal{O}(\varepsilon^2)$, $\mathbf{G}(\boldsymbol{q}_0^\varepsilon)\dot{\boldsymbol{q}}_0^\varepsilon = \mathcal{O}(\varepsilon^2)$ there exist initial values $(\boldsymbol{q}(t_0), \dot{\boldsymbol{q}}(t_0))^\top \in \mathfrak{M}^{\mathrm{s}} \times T_{\boldsymbol{q}(t_0)}\mathfrak{M}^{\mathrm{s}}$ with differences $\boldsymbol{q}(t_0) - \boldsymbol{q}_0^\varepsilon$, $\dot{\boldsymbol{q}}(t_0) - \dot{\boldsymbol{q}}_0^\varepsilon$ situated in the $\mathbf{M}(\boldsymbol{q}(t_0))$-orthogonal complement of the tangential space of $\mathfrak{M}^{\mathrm{s}}$ in $\boldsymbol{q}(t_0)$ such that for $\varepsilon < C_0 h$ and $C_0 > 0$ sufficiently small, the numerical approximations of (4.2) satisfy

$$\begin{aligned} \|\boldsymbol{q}(t_n) - \boldsymbol{q}_n^\varepsilon\| + \|\mathbf{P}_n(\dot{\boldsymbol{q}}(t_n) - \boldsymbol{v}_n^\varepsilon)\| &\le C\left(\varepsilon^2 + h^2\right) , \\ \|\boldsymbol{q}(t_n) - \boldsymbol{q}_n^\varepsilon\| + \|\dot{\boldsymbol{q}}(t_n) - \boldsymbol{v}_n^\varepsilon\| &\le C\left(\varepsilon^2 + h^2 + \varrho^n h\right) , \end{aligned}$$

whenever $t_n = t_0 + nh \in [t_0, t_{\mathrm{end}}]$. The constant $C > 0$ is independent of h, ε and n and $\boldsymbol{q}(t)$, $t \in [t_0, t_{\mathrm{end}}]$, denotes the solution of the DAE system (2.12). The constant $\varrho \in [\varrho_\infty, 1)$ also depends on the ratio $\frac{\varepsilon}{h}$. For the Chung–Hulbert(ϱ_∞) algorithm and $\varepsilon \ll h$ it can be chosen arbitrarily close to ϱ_∞.

Proof of Theorem 5.40. We use exactly the same techniques as in the previous proofs in the DAE and strongly damped SPP case: Collecting the estimates from Corollaries 5.5 and 5.38 and Lemma 5.36 and introducing the condensed error terms

$$\tilde{\boldsymbol{E}}_n^{\boldsymbol{y}} := \begin{pmatrix} \boldsymbol{e}_n^{\boldsymbol{q}} & \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{v}} \end{pmatrix}^\top , \quad \tilde{\boldsymbol{E}}_n^{\boldsymbol{z}} := \begin{pmatrix} \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} & h\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} & \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} & h\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} \end{pmatrix}^\top .$$

allow for an application of Lemma 5.1. In contrast to the proof in the strongly damped case it is not possible to eliminate the mass-orthogonal error components on velocity level $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}$ and they are included in the "algebraic" error terms $\boldsymbol{E}_n^{\boldsymbol{z}}$. The convergence result as well as the the numerical tests in Chapters 1 and 6 nevertheless reveal a reduction to only first order of convergence in this term and motivate this change. The overall error amplification is very similar to the one from the strongly damped case. Merely the lower two-by-two block in (5.35) is to be replaced by

$$\begin{aligned} &\overbrace{\begin{pmatrix} \frac{\varepsilon^2}{h^2}\mathbf{S}_n^{-1} & \mathbf{0} & -\beta\mathbf{I} \\ \mathbf{0} & \mathbf{I} & -\gamma\mathbf{I} \\ (1-\alpha_f)\mathbf{I} & \mathbf{0} & (1-\alpha_m)\mathbf{I} \end{pmatrix}}^{=\tilde{\mathbf{T}}_{1,\varepsilon}(\varepsilon^2/h^2\mathbf{S}_n^{-1})} \begin{pmatrix} h\boldsymbol{e}_{n+1}^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} \\ \boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{v}} \\ h\boldsymbol{e}_{n+1}^{\mathbf{G}\boldsymbol{a}} \end{pmatrix} = \overbrace{\begin{pmatrix} \frac{\varepsilon^2}{h^2}\mathbf{S}_n^{-1} & \mathbf{I} & (0.5-\beta)\mathbf{I} \\ \mathbf{0} & \mathbf{I} & (1-\gamma)\mathbf{I} \\ -\alpha_f\mathbf{I} & \mathbf{0} & -\alpha_m\mathbf{I} \end{pmatrix}}^{=\tilde{\mathbf{T}}_{2,\varepsilon}(\varepsilon^2/h^2\mathbf{S}_n^{-1})} \begin{pmatrix} h\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\varepsilon} \\ \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}} \\ h\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}} \end{pmatrix} \\ &\qquad + \mathcal{O}(h^2) + \mathcal{O}(\varepsilon^2) + \mathcal{O}(1)\boldsymbol{\eta}_{n,n+1}^{(\mathrm{SPP}),\varepsilon} , \end{aligned}$$

where the inverse of the Delassus matrix $\mathbf{S}_n := \mathbf{S}(\boldsymbol{q}(t_n))$ is again included in the leading linear propagation. It introduces a perturbation to the index-3 propagation matrix, see (5.22). As in the proof of Theorem 5.32 the matrix $\mathbf{S}_n^{-1}$ can be decomposed like in (5.36) such that in the new error terms

$$\boldsymbol{E}_n^{\boldsymbol{y}} := \tilde{\boldsymbol{E}}_n^{\boldsymbol{y}} , \quad \boldsymbol{E}_n^{\boldsymbol{z}} := \begin{pmatrix} \boldsymbol{e}_n^{\mathbf{P}\boldsymbol{a}} & h\boldsymbol{e}_n^{\mathbf{U}^\top\mathbf{S}\boldsymbol{\lambda},\varepsilon} & \boldsymbol{e}_n^{\mathbf{U}^\top\mathbf{G}\boldsymbol{v}} & h\boldsymbol{e}_n^{\mathbf{U}^\top\mathbf{G}\boldsymbol{a}} \end{pmatrix}^\top$$

we arrive at $n_{\boldsymbol{\lambda}}$ decoupled three-dimensional recursions of the above form for the last $3n_{\boldsymbol{\lambda}}$ components in $\boldsymbol{E}_n^{\boldsymbol{z}}$. As we have already intensively studied the eigenvalues of these matrices in Chapter 4 (and even defined stable Newmark methods in that way) we get contractivity by the same reasoning as in the strongly damped case and the assertion follows from Corollary 5.2 and $\mathbf{T} := \operatorname{blkdiag}\left(\frac{-\alpha_m}{1-\alpha_m}\mathbf{I}, [\tilde{\mathbf{T}}_{1,\varepsilon}^{-1}\tilde{\mathbf{T}}_{2,\varepsilon}](\frac{\varepsilon^2}{h^2}\mathbf{S}_n^{-1})\right)$. Contractivity for the first $n_{\boldsymbol{\lambda}}$ components in $\boldsymbol{E}_n^{\boldsymbol{z}}$ is ensured by the ODE zero stability condition $\alpha_m < \frac{1}{2}$ again. □

Note that the only requirements concerning the stability of the methods (in all four cases: index-2/3 and damped and stiff SPP) was the *linear* stability for the integration of the harmonic oscillator (4.12) or its strongly damped counterpart (5.26). We did not impose any nonlinear stability notion. Erlicher et al. (2002) found that nonlinear stability (in terms of G-stability or energy-stability) for Newmark-type integration methods cannot be shown in general anyway.

In Chapter 4 we saw that in the limit of infinite stiffness the Jordan decomposition of the amplification matrices may degenerate such that a diagonalization of the matrix is no longer possible. However, it is still possible to explicitly construct a suitable matrix norm in this case as we will show in the following example.

Example 5.42 (Construction of a suitable norm for the Chung–Hulbert(ϱ_∞) algorithm)
In case of the parameter set (4.4) of 'the' generalized-α algorithm the error amplification matrix

$$\tilde{\mathbf{T}}(\frac{\varepsilon^2}{h^2}\mathbf{S}_n^{-1};\varrho_\infty) := \tilde{\mathbf{T}}_{1,\varepsilon}^{-1}(\frac{\varepsilon^2}{h^2}\mathbf{S}_n^{-1})\tilde{\mathbf{T}}_{2,\varepsilon}(\frac{\varepsilon^2}{h^2}\mathbf{S}_n^{-1}),$$

only depends on the user-defined numerical damping parameter $0 \leq \varrho_\infty < 1$. Its Jordan canonical form for $n_{\boldsymbol{q}} = 1$ is given by as for the matrix $\mathbf{T}(\infty)$ in the linear case, see Section 4.2.2.

$$\tilde{\mathbf{T}}(0;\varrho_\infty) = \tilde{\mathbf{C}} \underbrace{\begin{pmatrix} -\varrho_\infty & 1 & 0 \\ 0 & -\varrho_\infty & 1 \\ 0 & 0 & -\varrho_\infty \end{pmatrix}}_{=\mathbf{J}} \tilde{\mathbf{C}}^{-1} \quad \text{introducing } \tilde{\mathbf{C}} := \begin{pmatrix} 1-\varrho_\infty^2 & \varrho_\infty - 2 & 0 \\ 0 & \frac{1-\varrho_\infty}{2(1+\varrho_\infty)} & \frac{-1}{(1+\varrho_\infty)^2} \\ 0 & 1 & 0 \end{pmatrix}.$$

To bound the norm from above we introduce the regular scaling matrix $\mathbf{D}_3 := \operatorname{diag}(1,\kappa,\kappa^2)$ for a (small) parameter $0 < |\kappa| < 1$. The subscript '3' shall relate to the three-by-three structure of the degenerate canonical form. A similarity transformation

$$\mathbf{D}_3^{-1}\mathbf{J}\mathbf{D}_3 = (\mathbf{C}\mathbf{D}_3)^{-1}\cdot\tilde{\mathbf{T}}(0,\varrho_\infty)(\mathbf{C}\mathbf{D}_3) = \begin{pmatrix} -\varrho_\infty & \kappa & 0 \\ 0 & -\varrho_\infty & \kappa \\ 0 & 0 & -\varrho_\infty \end{pmatrix}, \tag{5.42}$$

reveals an infinity norm $\|\bullet\|_\infty$ arbitrarily close to ϱ_∞, i. e.,

$$\|\bullet\|_\circ = \|(\tilde{\mathbf{C}}\mathbf{D}_3)^{-1}\cdot\bullet\cdot(\tilde{\mathbf{C}}\mathbf{D}_3)\|_\infty .$$

As a result, an appropriate vector norm is given by $\|\bullet\|_\circ := \|(\tilde{\mathbf{C}}\mathbf{D}_3)^{-1}\bullet\|_\infty$.

If in (5.42) the scaling coefficient $\kappa := \frac{1-\varrho_\infty}{2}$ is introduced, one can explicitly calculate the bound for $(\varepsilon^2/h^2)s^{-1}$ marking where the estimate $\|\tilde{\mathbf{T}}(\varepsilon^2/h^2 s^{-1};\varrho_\infty)\|_\circ < 1$ remains valid. It is given by

$$\frac{\varepsilon^2}{h^2}s^{-1} < \begin{cases} \dfrac{1-2\varrho_\infty+\varrho_\infty^2}{2(1+\varrho_\infty)^2(3+13\varrho_\infty+4\varrho_\infty^2)} & \text{if } \varrho_\infty \leq \frac{\sqrt{241}-13}{18} \\ \dfrac{1-\varrho_\infty}{10(1+\varrho_\infty)^3} & \text{otherwise} \end{cases}$$

and sketched (with a numerical calculation of the norm $\|\bullet\|_\circ$) in Figure 5.6. Note the different color scaling which ranges only from 0.5 to 1.0 as compared to Figure 5.5.

For other parameter choices as HHT(α), WBZ(α) or the algorithm with 'improved transient behavior' Gen(ϱ_∞,ϕ_0) from Remark 4.21 the Jordan canonical form in the limit case $\varepsilon \to 0$ has different forms as presented in (4.32). In these cases, a scaling using the matrices $\mathbf{D}_1 := \mathbf{I}$ or $\mathbf{D}_2 := \operatorname{diag}(1,\kappa,1)$ can be utilized to receive boundedness. ♢

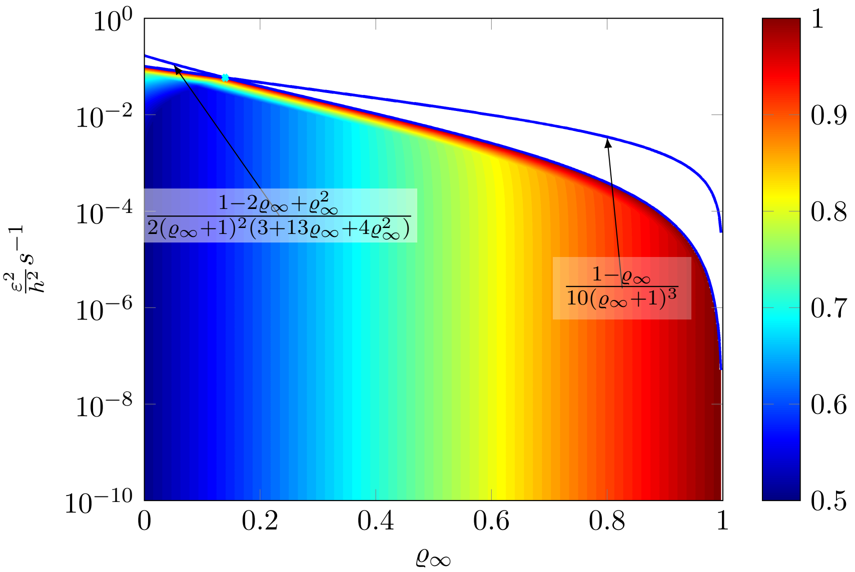


Figure 5.6: $\|\bullet\|_\circ$-norm of lower three-by-three block of the amplification matrix in the stiff cases

5.4 Summary

Remark 5.43 (Limitations of the convergence results)
To obtain the above result, we explicitly needed both: Assumptions 5.3 and 5.25 and if any of them is violated, the results do not remain valid any longer.

In particular, if we ignore Assumption 5.3 and reduce the time step size h more and more, at a certain point, we reach the area of classical convergence theory (Erlicher et al., 2002). In this case, the ε^2 bound for the errors on velocity level is too optimistic and needs to be adapted to linear convergence in ε. The reason is the 'modeling error' of the singularly perturbed models when compared to the solution of the constrained mechanical system (2.12): Equation (3.22) that on velocity level the difference fulfills

$$\mathbf{G}(\boldsymbol{q}(t_n))(\boldsymbol{q}^\varepsilon(t_n)-\boldsymbol{q}(t_n))=\mathcal{O}(\varepsilon) \quad \Rightarrow \quad \boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}=\mathbf{G}(\boldsymbol{q}(t_n))(\boldsymbol{q}_n^\varepsilon-\boldsymbol{q}^\varepsilon(t_n)+\underbrace{\boldsymbol{q}^\varepsilon(t_n)-\boldsymbol{q}(t_n)}_{=\mathcal{O}(\varepsilon)}).$$

Practically, Theorem 5.40 therefore only applies for the error terms from $\boldsymbol{E}_n^{\boldsymbol{y}}$ because $\boldsymbol{q}^\varepsilon$ and $\mathbf{P}\dot{\boldsymbol{q}}^\varepsilon$ remain ε^2-close to the DAE solution, see Corollary 3.18. In short, the main result from Theorem 5.40 is the order reduction on velocity level and numerical experiments can only verify the estimate

$$\|\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}\|\leq C\left(h^2+h\varrho^n+\varepsilon\right).$$

If we were, on the other hand, to disregard Assumption 5.25 the influence of initial error terms would become too severe and convergence to $(\boldsymbol{q}(t),\dot{\boldsymbol{q}}(t))^\top$, defined by mass-orthogonal projection of initial values, can no longer be shown. Numerical tests, nevertheless, indicate that this is no stability problem of the time integration itself but rather that some sort of phase shift is introduced, see the discussion in Chapter 6.

Note that the scaling of the global error terms $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{a}}$ and $\boldsymbol{e}_n^{\mathbf{S}\boldsymbol{\lambda},\varepsilon|\delta}$ by the time step size h has only been used to cope with the large initial errors. Once the algorithm exits the initial phase, the error analysis could be adapted such that first order error terms are damped out and become negligible after a transient phase.

At last, note also that for the contractivity result one key ingredient was norm equivalence in $\mathbb{R}^n$. An extension to PDE models (especially those of structural dynamics, see Example 3.19) of mathematically very similar structure as investigated by Altmann (2015) requires different techniques.

Remark 5.44 (Scaling of the error components)
It is as well possible to consider a different scaling of the error components to be 'consistent' with the classical analysis of Chung and Hulbert (1993) of the harmonic oscillator (4.12). *In that case one would have to use* continuity at complex infinity *as it has been done, for example, by Schneider (1995) and the amplification matrix would resemble the one in* (4.15)*, see also Remark 4.6.*

Remark 5.45 (Mathematical pendulum revisited)
With the comprehensive error analysis of this chapter at hand, we are now prepared to explain all observations from the introductory Chapter 1.

The fact that the errors in the transient phase of the time integration may grow is due to the overshoot phenomenon already discussed in Remark 4.21 and Section 5.2. It carries over from the index-3 case to stiff mechanical systems as their governing linear error amplification is alike. An adaptation of the parameters (HHT, WBZ, Gen) may lower (yet not completely avoid) the outcome of overshoot. We will demonstrate this in Chapter 6 below.

The analysis revealed that convergence for position and velocity components is ensured as long as Assumptions 5.3 and 5.25 are fulfilled; the pendulum example already shows that the bounds on the initial deviations from the constraint manifolds $\mathfrak{M}^{\mathrm{s}}$, $\mathfrak{M}^{\mathrm{d}}$ respectively, are sharp, a deviation of ε^2 or δ is possible, larger initial errors result in divergence. The qualitative behavior of the global errors in Figure 1.3 are in perfect agreement with the error estimates from Corollaries 5.14 and 5.22. Even the ε^{-2} and δ^{-1} behavior in the diverging case can be explained by Lemmas 5.31 and 5.39 as the initial errors on level of acceleration-like variables and Lagrange multipliers have this form.

Second order of convergence from the classical setting of nonstiff ODEs carries over to the DAE and SPP setting only in the index-2 and strongly damped case. For the index-3 formulation we showed ways to overcome reduction in order of convergence for the tangential velocity error terms $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}$ in Remark 5.23; for singularly perturbed systems such a modification of initial values might also be possible but does not seem to be practical having in mind that for most applications one is faced with SPPs where the underlying constraint equations are unknown anyway.

Chapter 6

Implementation and numerical tests

The following chapter is devoted to practical aspects of Newmark integrators (4.2) in the context of singularly perturbed or differential-algebraic systems. We start with a brief discussion of the algorithm from an application perspective in Section 6.1 putting emphasis on a reliable realization of the corrector iteration that is not too strongly affected by the SPP nature of the problems and on certain details as scaling and the choice of the damping parameter in the Newton–Raphson method. In Section 6.2 we give an overview on the used benchmark problems before the theoretical findings from the previous chapters are verified in Sections 6.3, 6.4 and 6.5. As we have already seen that order reduction may occur and rather strong assumptions need to be fulfilled for the SPP problems to apply the above convergence results, we will give an outlook on possible improvements for the initialization of the algorithm in Section 6.6.

6.1 Implementation aspects

Before the numerical properties and convergence results established in Chapters 4 and 5 are discussed by numerical tests for benchmark problems from the literature, we point out a series of details concerning the practical viewpoint of implementation: It is well-known that the most severe challenge of numerical integrators for singularly perturbed problems using moderate time step sizes is the robust and stable solution of the corrector equations.

In Section 6.1.1 we present a solution procedure for (4.2) that, in accordance with the requirements from Remark 4.19, allows for a onestep representation and only involves the solution of linear systems of dimension $n_{\boldsymbol{q}}$ for ODE systems or $n_{\boldsymbol{q}} + n_{\boldsymbol{\lambda}}$ for DAE systems, respectively. The latter one is also favorable for the solution of SPP systems of stiff or strongly damped type if considered in their index-1 forms (3.7) or (3.16) because the convergence of the corrector iterations is guaranteed for time step sizes h that are independent of the penalty parameter. Unfortunately, in concrete simulation environments it is often not possible to simply reformulate the equations of motion to a numerically more favorable form. Sometimes fundamental terms as the constraint Jacobian matrix $\mathbf{G} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{q}}$ or even the mass matrix $\mathbf{M}$ are not possible to obtain from a (black-box) multibody or FE-tool.

6.1.1 Onestep solution procedure

The second order nature of the equations of motion of mechanical systems always allows for drastic savings within the numerical solution because the relation $\boldsymbol{v} = \dot{\boldsymbol{q}} = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{q}$ of generalized position and velocity coordinates is linear, a fact which can explicitly be used. For algorithms that are designed for second order ODE/DAE systems as the Newmark family (4.2) this is naturally reflected by the linear structure of (4.2a) and (4.2b). As a result, the corrector equations,

i. e., the numerical solution procedure to obtain approximations in the current time step, can be implemented in a way that the linear systems that occur in the iterative solution of the nonlinear systems are only of dimension $n_{\boldsymbol{q}}(\,+\,n_{\boldsymbol{\lambda}})$. The following algorithm which is adapted from the ones given by Géradin and Cardona (2001) and Arnold and Brüls (2007) meets this requirement.

```
Algorithm: $(\boldsymbol{q}_{n+1}, \boldsymbol{v}_{n+1}, \dot{\boldsymbol{v}}_{n+1}, \boldsymbol{\lambda}_{n+1}, \boldsymbol{a}_{n+1}) = \mathrm{NewmarkStep}(\boldsymbol{q}_n, \boldsymbol{v}_n, \dot{\boldsymbol{v}}_n, \boldsymbol{\lambda}_n, \boldsymbol{a}_n)$

Given the ODE/DAE system (4.1);
/* Use initial guesses for acceleration variables and lambda                 */
$\boldsymbol{\lambda}_{n+1} := \boldsymbol{\lambda}_{n+1}^{(0)}$;
$\dot{\boldsymbol{v}}_{n+1} := \dot{\boldsymbol{v}}_{n+1}^{(0)}$;
/* Initial guesses:  acceleration-like, position and velocity variables   */
$\boldsymbol{a}_{n+1} := \frac{1}{1-\alpha_m}((1-\alpha_f)\dot{\boldsymbol{v}}_{n+1} + \alpha_f \dot{\boldsymbol{v}}_n - \alpha_m \boldsymbol{a}_n)$;
$\boldsymbol{q}_{n+1} := \boldsymbol{q}_n + h\boldsymbol{v}_n + h^2(0.5-\beta)\boldsymbol{a}_n + h^2\beta\boldsymbol{a}_{n+1}$;
$\boldsymbol{v}_{n+1} := \boldsymbol{v}_n + h(1-\gamma)\boldsymbol{a}_n + h\gamma\boldsymbol{a}_{n+1}$;
/* Corrector iteration                                                        */
for $i = 1 : i_{\max}$ do
    Calculate the residuals in $\mathbf{M}\dot{\boldsymbol{v}} - \boldsymbol{F} = \mathbf{0}$, $(\boldsymbol{r_q})$ and $\boldsymbol{\Phi} = \mathbf{0}$, $(\boldsymbol{r_\lambda})$;
    if $\|(\boldsymbol{r_q}, \boldsymbol{r_\lambda})^\top\|^* < Tol_{\mathrm{Newton}}$ then
        break;
    $\boldsymbol{d} := -\mathbf{S}_t^{-1}\begin{pmatrix}\boldsymbol{r_q}\\ \boldsymbol{r_\lambda}\end{pmatrix}$;
    find $\sigma > 0$ using a line search strategy;
    $\begin{pmatrix}\boldsymbol{\Delta q}\\ \boldsymbol{\Delta\lambda}\end{pmatrix} := \sigma\boldsymbol{d}$;
    $\boldsymbol{q}_{n+1} := \boldsymbol{q}_{n+1} + \boldsymbol{\Delta q}$;
    $\boldsymbol{v}_{n+1} := \boldsymbol{v}_{n+1} + \gamma'\boldsymbol{\Delta q}$;
    $\dot{\boldsymbol{v}}_{n+1} := \dot{\boldsymbol{v}}_{n+1} + \beta'\boldsymbol{\Delta q}$;
    $\boldsymbol{a}_{n+1} := \boldsymbol{a}_{n+1} + \bar{\beta}\boldsymbol{\Delta q}$;
    $\boldsymbol{\lambda}_{n+1} := \boldsymbol{\lambda}_{n+1} + \Delta\boldsymbol{\lambda}$;
```

For ODE systems, including SPPs, the algorithm is formally obtained for $n_{\boldsymbol{\lambda}} = 0$, there is no additional adaptation necessary. The very large stiffness in the SPP cases, however, reduces the range of convergence for the Newton–Raphson method such that the corrector iteration may only succeed for very small time step sizes.

The integer value $i_{\max} > 0$ is the maximum number of Newton–Raphson iterations and typically set to about twenty. For real-world applications it usually suffices to perform less than five iterations, for higher precision requirements this number may increase (Eich-Soellner and Führer, 1998, Arnold et al., 2011). The tolerance of the corrector iteration $Tol_{\mathrm{Newton}} > 0$ is usually adapted to the user-defined bounds for the global error of the time integration method. The specific norm $\|\cdot\|^*$ typically includes a proper scaling of the involved residual terms and may depend on the formulation, see Remark 6.2 below. Deuflhard (2004) favors the so-called 'affine invariant version' of the stopping criterion that only involves the norm of the increments $(\Delta\boldsymbol{q}, \Delta\boldsymbol{\lambda})^\top$ instead of the residuals, see related current results of Arnold and Hante (2016) for the Chung–Hulbert parameter set. In a practical implementation one usually combines absolute and relative errors when checking the stopping criterion. For most of the benchmark tests below the iterations have been carried out 'until convergence', i. e., until a very low tolerance ($Tol_{\mathrm{Newton}} \approx 10^{-30}$; in these cases we did the calculations with 100 digit precision) has been reached.

The involved system matrices $\mathbf{S}_t$ for the DAE and SPP cases as well as the 'line search strategy' in line 12 are discussed below. The derived parameter values β', γ' and $\bar{\beta}$ are defined as

$$\bar{\beta} := \frac{1}{h^2\beta}, \quad \beta' := \frac{(1-\alpha_m)\bar{\beta}}{1-\alpha_f} = \frac{1-\alpha_m}{h^2\beta(1-\alpha_f)}, \quad \gamma' := \gamma h\bar{\beta} = \frac{\gamma}{h\beta}\,.$$

The reduction of the dimension can be seen as a quasi-standard for commercial codes in mechanical system simulation. Naturally, it is also possible to use larger system matrices $\mathbf{S}_t$ that include the iteration of the variables $\boldsymbol{q}$, $\boldsymbol{v}$, $\boldsymbol{a}$ and $\boldsymbol{\lambda}$ 'although [this is] probably never employed in practice' (Bottasso et al., 2008).

Remark 6.1 (Interpretation of the condensed Newton–Raphson iteration)

(a) The procedure of the above algorithm is in the engineer's literature known as 'static condensation'. To motivate the procedure Géradin and Cardona (2001) interpret the derived parameters

$$\frac{\partial\dot{\boldsymbol{v}}_{n+1}}{\partial\boldsymbol{q}_{n+1}} = \beta'\mathbf{I}\,, \quad \frac{\partial\boldsymbol{v}_{n+1}}{\partial\boldsymbol{q}_{n+1}} = \gamma'\mathbf{I}\,, \quad \frac{\partial\boldsymbol{a}_{n+1}}{\partial\boldsymbol{q}_{n+1}} = \bar{\beta}\mathbf{I}$$

as coefficients of internal differentiations such that the overall algorithm is a simplification of Newton's method for the entire $(4n_{\boldsymbol{q}} + n_{\boldsymbol{\lambda}})$-system to obtain $\boldsymbol{q}_{n+1}$, $\boldsymbol{v}_{n+1}$, $\boldsymbol{a}_{n+1}$, $\dot{\boldsymbol{v}}_{n+1}$ and $\boldsymbol{\lambda}_{n+1}$ in the sense that just the chain rule is used for the linearization of the equations.

(b) Any general coupled system of linear and nonlinear equations of the form

$$\mathbf{0} = \Xi(\boldsymbol{y},\boldsymbol{z}) := \begin{pmatrix} \mathbf{A}_1\boldsymbol{z} - (\mathbf{A}_2\boldsymbol{y} + \boldsymbol{b}) \\ \tilde{\Xi}(\boldsymbol{y},\boldsymbol{z}) \end{pmatrix}\,, \quad (\Xi \in \mathbb{R}^{n_{\boldsymbol{y}}+n_{\boldsymbol{z}}},\, \tilde{\Xi} \in \mathbb{R}^{n_{\boldsymbol{y}}},\, \det(\mathbf{A}_1) \neq 0)\,, \tag{6.1}$$

may equivalently be stated as condensed system

$$\mathbf{0} = \Xi^{\text{cond.}}(\boldsymbol{y}) := \tilde{\Xi}(\boldsymbol{y}, \mathbf{A}_1^{-1}(\mathbf{A}_2\boldsymbol{y} + \boldsymbol{b})) \in \mathbb{R}^{n_{\boldsymbol{y}}}$$

by eliminating the linear constraints in Ξ. In the present situation we may choose the variables $\boldsymbol{z} := (\boldsymbol{v}_{n+1}, \dot{\boldsymbol{v}}_{n+1}, \boldsymbol{a}_{n+1})^\top$ and $\boldsymbol{y} := (\boldsymbol{q}_{n+1} - \boldsymbol{q}_n, \boldsymbol{\lambda}_{n+1})^\top$. The nonlinear system $\mathbf{0} = \Xi$ comprises the linear equations (4.2a), (4.2b) *and* (4.2c) *forming the upper block and $\tilde{\Xi}$ realized by the nonlinear dynamic equations and constraints* (4.2d). *The relation $\boldsymbol{z} = \mathbf{A}_1^{-1}(\mathbf{A}_2\boldsymbol{y} + \boldsymbol{b})$ reads in detail*

$$\begin{aligned}
\boldsymbol{v}_{n+1} &= (1 - h\gamma')\boldsymbol{v}_n + (1 - h\gamma'/2)h\boldsymbol{a}_n + \gamma'\boldsymbol{\Delta q}\,,\\
\dot{\boldsymbol{v}}_{n+1} &= -h\beta'\boldsymbol{v}_n + (1/(1-\alpha_f) - (h^2\beta')/2)\boldsymbol{a}_n - \alpha_f/(1-\alpha_f)\dot{\boldsymbol{v}}_n + \beta'\boldsymbol{\Delta q}\,,\\
\boldsymbol{a}_{n+1} &= -h\bar{\beta}\boldsymbol{v}_n + (1 - h^2\bar{\beta}/2)\boldsymbol{a}_n + \bar{\beta}\boldsymbol{\Delta q}\,,
\end{aligned}$$

with $\boldsymbol{\Delta q} = \boldsymbol{q}_{n+1} - \boldsymbol{q}_n$. It can easily be seen that the Newton–Raphson iteration for Ξ leads to the same approximations as the one for $\Xi^{\text{cond.}}$ if the system matrices $\mathbf{S}_t$ are defined as stated below and in the latter one the iterates for $\boldsymbol{z}$ are defined by $\boldsymbol{z}^{(k+1)} := \mathbf{A}_1^{-1}(\mathbf{A}_2\boldsymbol{y}^{(k+1)} + \boldsymbol{b})$, $k = 0, 1, \ldots$, which is the purpose of lines 14 to 18 in the above algorithm.

(c) One might also interpret the separation of linear and nonlinear equations as a block Gauss elimination *as Lubich (1991) uses it for a linearly implicit Euler method (with extrapolation) in the multibody code MEXX, see also (Lubich et al., 1992).*

In the literature (cf. Arnold and Brüls, 2007) the update of acceleration-like variables $\boldsymbol{a}_{n+1}$ is sometimes excluded from the (inner) for-loop which slightly improves the efficiency of the algorithm. Note that, apart (maybe) from the initialization of $\boldsymbol{\lambda}_{n+1}^{(0)}$, the input variable $\boldsymbol{\lambda}_n$ does not affect the algorithm. This reflects the Hessenberg-structure of the system: It is said that the equations/numerical procedure possess 'no memory' in $\boldsymbol{\lambda}$. In the convergence analysis this is a crucial, yet somewhat hidden, part: The lower block in the global error recursion is contractive and not of the form $\mathbf{I} + \mathcal{O}(h) + \mathcal{O}(\varepsilon^2)$ as it would be for nonstiff ODEs (Erlicher et al., 2002). So, lower order in the local error does not affect the overall error bound (at least not too much), cf. Examples 5.28 and 5.35. In the benchmarks below $\boldsymbol{\lambda}_{n+1}^{(0)} := \mathbf{0}$ has been used in accordance to the original pseudocode of Arnold and Brüls (2007).

Initial values for the corrector iteration Simply inserting the linear equations to lower the dimension of the system as proposed in the algorithm 'NewmarkStep' is not enough. Static condensation also requires that the initial values already fulfill the linear equations. In the above pseudocode this requirement has influence on the initializations of position, velocity, and acceleration-like variables in lines 4, 5 and 6. These are constructed such that (4.2a), (4.2b) and (4.2c) are fulfilled before the variables enter the iteration.

On position and velocity level the above procedure may be interpreted as truncated forward Euler steps which are corrected using the estimate $\boldsymbol{a}_{n+1}$. This is evidently not the only way to choose these values. Jansen et al. (2000) present numerical experiments using different initializations, among them zero-acceleration/velocity, constant extrapolation, linear extrapolation, and even a backward Euler predictor-step. Exhaustive numerical experiments lead to the suggestion that linear extrapolation should be used (in the context of the simulation of filtered Navier–Stokes equations). For the benchmarks below we used the initial guesses of the above algorithm nonetheless because in the engineer's literature (Géradin and Cardona, 2001) the above 'forward Euler' predictors are typically preferred together with a zero-acceleration condition

$$\dot{\boldsymbol{v}}_{n+1}^{(0)} := \mathbf{0}\,,$$

which is motivated by static analysis of structures and performs often more robust for numerical experiments with immense overshoot. We have found that, especially after the transient phase and for high precision demands, the initialization

$$\dot{\boldsymbol{v}}_{n+1}^{(0)} := \frac{\alpha_m \boldsymbol{a}_n - \alpha_f \dot{\boldsymbol{v}}_n}{\alpha_m - \alpha_f}\,, \quad (\text{if } \alpha_m - \alpha_f \neq 0)$$

often decreases the required number of Newton–Raphson iterations. The idea behind this construction is that then $\dot{\boldsymbol{v}}_{n+1}^{(0)} = \boldsymbol{a}_{n+1}$ holds after line 4 which seems reasonable as both values approximate roughly the same quantity.

System matrices The Jacobian in the Newton–Raphson iteration needs to take into account static condensation as well. For the index-3 integrator it is given by:

$$\mathbf{S}_t := \begin{pmatrix} \beta'\mathbf{M} + \gamma'\mathbf{C}_t + \mathbf{K}_t & \mathbf{G}^\top \\ \mathbf{G} & \mathbf{0} \end{pmatrix},$$

for the index-2 integrator it should be defined as

$$\mathbf{S}_t := \begin{pmatrix} \beta'\mathbf{M} + \gamma'\mathbf{C}_t + \mathbf{K}_t & \mathbf{G}^\top \\ \gamma'\mathbf{G} + \mathbf{Z}_t & \mathbf{0} \end{pmatrix}.$$

For the index-1 formulations (3.7) and (3.16) of the SPP problems, i. e., (5.25) and (5.39) b), the zero matrices need to be replaced by $-\delta\mathbf{I}$ or $-\varepsilon^2\mathbf{I}$ since here the constraint equations explicitly depend on the (artificial) Lagrange multipliers $\boldsymbol{\lambda}^{\varepsilon|\delta}(t)$. The involved matrices $\bullet_t$ are defined by $\mathbf{K}_t{:=}\partial(\mathbf{M}\dot{\boldsymbol{v}}{-}\boldsymbol{F})/\partial\boldsymbol{q}$, which is denoted as (tangent) stiffness matrix, $\mathbf{C}_t{:=}\partial(-\boldsymbol{F})/\partial\dot{\boldsymbol{q}}$ representing the (tangent) damping matrix, and $\mathbf{Z}_t{:=}\mathsf{R}(\boldsymbol{q})(\boldsymbol{v},\cdot)$, the (tangent) constraint curvature matrix. In case of the stabilized index-2 formulation, see Remark 5.18, one more set of nonlinear equations and the additional variables $\boldsymbol{\mu}_n$ are appended to the system, see (Arnold et al., 2016). In this case it is possible to split the computation into the solution of two linear systems of dimension $n_{\boldsymbol{q}} + n_{\boldsymbol{\lambda}}$ which may save costs even further.

In the benchmarks below the system matrices have been calculated analytically. For most real-world problems such a procedure is not possible without large computational effort and/or the aid of automatic differentiation routines. Due to the black-box character of many substructures and the use of surrogate models without sufficient smoothness the Jacobians have to be approximated numerically. The high cost of such approximations is then counterbalanced by the usage of a simplified Newton–Raphson method, i. e., through keeping the Jacobian constant over multiple time steps. The reevaluation of the system matrices is typically based on heuristics that involve the behavior of the solution and the asymptotic behavior of the iteration which is only linear for the simplified Newton–Raphson method (Eich-Soellner and Führer, 1998). For BDF methods in multibody dynamics and without condensation of the nonlinear systems, Arnold et al. (2011) report that one Jacobian evaluation per ten time steps is common.

In the same virtue of using only approximate Jacobians it is, of course, possible to leave certain submatrices out: Brüls (2005) favors to use $\mathbf{K}_t \approx -\partial\boldsymbol{F}/\partial\boldsymbol{q}$ instead of the above definition because for many multiphysics applications this matrix can be obtained rather cheap and often even analytically. Also the scaling behavior of the derived parameters is important. For small time step sizes it holds $\beta' \gg \gamma' \gg 1$ such that omitting the evaluation of $\mathbf{C}_t$ and $\mathbf{K}_t$ is often justified as well. This especially bears an advantage if the mass and constraint Jacobian matrix are known since then (in the index-3 and stiff case) no difference approximation is necessary at all. Also, in many cases $\mathbf{M}$ is sparse or easy to invert which saves even more computational effort. For SPPs in the ODE formulation, however, the very large stiffness parameters enter the tangent stiffness and/or damping matrix and a truncation weakens the convergence behavior of the iteration.

Remark 6.2 (Newton–Raphson variables)
In the classical setting of structural dynamics with its primarily linear or quasi-linear equations, the update formulae are usually expressed and implemented in terms of $\boldsymbol{\Delta a} := \boldsymbol{a}_{n+1} - \boldsymbol{a}_n$ rather than $\boldsymbol{\Delta q}$ (see for instance Hoff and Pahl, 1988a, Chung and Hulbert, 1993, Erlicher et al., 2002). A reformulation with different variables is easily made: The variables in (6.1) *enter $\boldsymbol{\Xi}$ in a (almost) symmetric way, such that the roles of components that enter $\boldsymbol{y}$ or $\boldsymbol{z}$ can be exchanged as long as the regularity of the corresponding matrix $\mathbf{A}_1$ is given. Negrut et al. (2005) argue that using $\boldsymbol{\Delta a}$ appears as the more natural choice since $\boldsymbol{\Delta a}$ and $\boldsymbol{\Delta\lambda}$ act on 'qualitatively the same kinematic level'. On the other hand, Hoff and Pahl (1988b) point out that an implementation with $\boldsymbol{\Delta q}$ is more convenient and natural since position variables are in fact in most cases the variables of actual interest. Negrut et al. (2005) argue that due to the structure of $\mathbf{S}_t$ the submatrices that usually have to be approximated by finite differences are multiplied by a factor in $\mathcal{O}(h^{-2})$ such that numerical errors are amplified; this is no longer the case for acceleration updates. In exact arithmetics (including a sufficiently accurate approximation of Jacobian matrices) the both formulations are, of course, equivalent.*

Remark 6.3 (Scaling of corrector equations: 'Balancing')
Static condensation cannot restrain the drawback that the linear systems are badly scaled (Petzold and Lötstedt, 1986) since the derived parameters $\beta'$and γ' enter the system matrix $\mathbf{S}_t$. Bottasso et al. (2008) propose a scaling strategy for balancing: Use the equivalence of two linear systems

$$\mathbf{S}_t \boldsymbol{d} = \boldsymbol{r} \quad \Leftrightarrow \quad \bar{\mathbf{S}}_t \bar{\boldsymbol{d}} = \bar{\boldsymbol{r}}\,, \tag{6.2}$$

where $\bar{\mathbf{S}}_t := \mathbf{L}\mathbf{S}_t\mathbf{R}$, So, instead of one inversion of $\mathbf{S}_t$, the right-hand side needs to be scaled like $\bar{\boldsymbol{r}} := \mathbf{L}\boldsymbol{r}$, and, after the solution of the right system in (6.2), $\boldsymbol{d} := \mathbf{R}\bar{\boldsymbol{d}}$ can be obtained. Matrices $\mathbf{L}$ and $\mathbf{R}$ are chosen such that the condition number $\bar{\mathbf{S}}_t$ is much less than that of $\mathbf{S}_t$ and that the multiplication with $\mathbf{L}$ and $\mathbf{R}$ is computationally cheap and the inversion of $\bar{\mathbf{S}}_t$ not drastically more expensive than for the original system (Petzold and Lötstedt, 1986). In practice one uses block diagonal matrices consisting of multiples of $\mathbf{I}$ such that matrix multiplication coincides with scalar multiplication.

For the index-2 systems the proposed 'optimal scaling' (in the sense of Bottasso et al. (2008)) is given by

$$\mathbf{R} := \mathrm{blkdiag}(\gamma\mathbf{I}, \tfrac{1}{\beta h^2}\mathbf{I})\,, \quad \mathbf{L} := \mathrm{blkdiag}(\beta h^2\mathbf{I}, h\mathbf{I})\,,$$

and for the index-3 system one obtains (see Arnold and Brüls, 2007)

$$\mathbf{R} := \mathrm{blkdiag}(\mathbf{I}, \tfrac{1}{\beta h^2}\mathbf{I})\,, \quad \mathbf{L} := \mathrm{blkdiag}(\beta h^2\mathbf{I}, \mathbf{I})\,.$$

The same scaling techniques can be applied for SPPs in their index-1 formulation.

Cardona and Géradin (1994) suggest to incorporate structural properties, i. e., to use an average mass/inertia term when scaling the equations. In all cases, from a practical point of view, it is a drawback that the scaling depends on the time step size which might cause more implementational inconvenience (Cardona and Géradin, 1994). In variable step size implementations for DAEs or stiff systems scaling of the residuals and using these scaled versions in the stopping criterion is common as it is a well-known result (Petzold, 1982) that the unstable perturbation behavior of high-index DAEs transfers to the numerical solution as well and hereby spoils step size control algorithms (Hairer et al., 1989a). Schaub and Simeon (2002) show that for stiff mechanical systems a similar scaling technique is necessary to gain efficiency. Yet, disabling the step size control or using fixed step size implementations for very stiff or high-index systems still seems to be the method of choice for most real-time simulations (Arnold et al., 2007) and is very popular for large scale simulations (Simeon, 1998) as well.

In summary, there is not 'one optimal way' of defining the variables in the Newton–Raphson iteration or scaling the equations or truncating the system matrices for all cases but one should instead adapt these issues to the problem at hand. The corrector iteration by means of Newton–Raphson iteration with static condensation is also at the heart of proving unique (local) existence of the numerical solutions of (4.2) in each time step which we will fix in the following Lemma 6.4. Its proof is skipped and can be found in (Arnold et al., 2016, Lemma 3.3) for the DAE case. For SPPs the adaptation is straightforward since the index-1 formulation with $\boldsymbol{\lambda}^{\delta|\varepsilon}$ is simply a perturbation of the DAE systems and the regularity of $\mathbf{S}_t$ is given as before. As the proof relies on a convergence result (Kelley, 1995) for the Newton–Raphson iteration in the above form its convergence is proven as well. For the original ODE formulation the existence and uniqueness does not imply that a 'standard Newton–Raphson iteration' converges (unless the time step size h is chosen unrealistically small or the penalizing potential/Rayleigh function is quadratic). Simeon (2013) uses a nonlinear extension of the Prothero–Robinson equation (5.40) to derive step size restrictions. A remedy for the bad convergence behavior without employing the index-1 form is given by Lubich (1993) for Runge–Kutta methods and was later on extended by Yen and Petzold (1998), Yen et al. (1998) to multistep integration methods, (CS-method, CM-method). From a theoretical point of view these methods are based on projection techniques and the tangent space parameterization introduced in Section 2.1.2.

Lemma 6.4 (Unique solvability)
There are constants $h_0, \gamma_0 > 0$ independent of δ or ε, respectively, such that for given initial values $(\boldsymbol{q}_0^{1|2|\delta|\varepsilon}, \boldsymbol{v}_0^{1|2|\delta|\varepsilon})^\top$ fulfilling

$$\begin{aligned} \|\boldsymbol{g}(\boldsymbol{q}_0^{2|3})\| &\le \gamma_0 h\,, & \|\mathbf{G}(\boldsymbol{q}_0^{2|3})\boldsymbol{v}_0^{2|3}\| &\le \gamma_0 & &\text{in the DAE case,} \\ \|\boldsymbol{g}(\boldsymbol{q}_0^{\delta})\| &\le \gamma_0 h\,, & \|\mathbf{G}(\boldsymbol{q}_0^{\delta})\boldsymbol{v}_0^{\delta}\| &\le \gamma_0\delta & &\text{for strongly damped systems,} \\ \|\boldsymbol{g}(\boldsymbol{q}_0^{\varepsilon})\| &\le \gamma_0 \varepsilon^2\,, & \|\mathbf{G}(\boldsymbol{q}_0^{\varepsilon})\boldsymbol{v}_0^{\varepsilon}\| &\le \gamma_0\varepsilon & &\text{for stiff mechanical systems} \end{aligned} \tag{6.3}$$

the algorithmic equations (4.2) (locally) possess a unique solution whenever $h \in (0, h_0]$ and $\gamma, \beta, 1-\alpha_m, 1-\alpha_f \neq 0$. The solutions fulfill qualitatively the same estimates as given in (6.3) such that induction leads to existence of numerical solutions on a finite time interval.

In the work of Stumpp (2004) and Lubich (1993) on numerical integration methods singular SPPs, the existence proofs involve the nonlinear coordinate transforms from Lemmas 3.7 and 3.14 which modify the singular terms to a linear form for which the existence proof is simpler. In the next remark we present techniques to improve the overall efficiency and robustness of the corrector iteration.

Remark 6.5 (Line search algorithms)
The goal of most line search algorithms for Newton-type iterations is to serve as a stabilization in the sense that robustness of the algorithm is improved while the good convergence in close proximity to the root is preserved. The main idea behind them is that damped Newton–Raphson iterations, i. e., those where the increment is multiplied by a constant factor $0 < \sigma < 1$, usually have a better global convergence behavior whereas the classical (possibly simplified) method is much faster in proximity of a solution. So, in each corrector step for a series of candidates σ of possible step sizes it is tested whether they fulfill a certain descent condition and if so, this length is used. The two most commonly used descent conditions are due to Wolfe (1969) and to Goldstein (1962) and Armijo (1966). The step size candidates are chosen by different heuristics; most common are the choice $\sigma = 1, \frac{1}{2}, \frac{1}{4}, \ldots$ or polynomial interpolation of the objective at the previous candidates. Numerical experiments suggest that approximately five candidates typically suffices in the context of ODE/DAE time integration methods. Especially after the transient phase the benefits from a line search method are mostly negligible as the algorithms only find step lengths $\sigma = 1$ anyway. For an overview on line search algorithms (in the context of optimization problems) we refer to (Kelley, 1995, Deuflhard, 2004).

Solving the nonlinear systems that arise in each time integration step using the index-1 formulation (as well as the CS/CM methods) is only possible if certain values of the DAE problem, most importantly the constraint Jacobian **G**, *can numerically be evaluated. As we have already explained in Chapter 3 this might be difficult or even impossible. In order to being able to tackle these problems also, one can employ* path-following methods. *Within this framework each time integration step is seen as dependent on a scalar parameter (mostly the step size h) and a step-by-step subroutine is used to follow the according solution branch. The crux is that these methods only make sense if the solution of all the auxiliary problems along the path are not computationally more expensive than using smaller time steps in the first place. This goal can for example be accomplished if the tolerances for the auxiliary problems is cruder than for the original problem. Also, one can use specialized step size control algorithms to improve efficiency. For further details we refer to Deuflhard (2004) and Rheinboldt (1980).*

But, nevertheless, the very limited success of those methods for industrial implementations or applications gives rise to the hypothesis that they are mostly of mathematical interest, see also the review article by Ascher et al. (2007). It is one of the main results of the work of Lubich (1993) that, if possible, one should always try to tackle the DAE problem, i. e., the underlying slow motion, directly.

6.2 Benchmarks

The general framework that determines the equations of motion in the case of strongly damped and stiff mechanical systems from Chapter 3 allows to give an alternative formulation of any given mechanical system of the form (2.12). Therefore, we describe the benchmarks in the following section only in their index-3 form as all alternative formulations can be derived from them. In the numerical experiments below we chose the benchmarks in a way that the phenomena to be shown are underlined well which is why we are going to switch between them from section to section.

6.2.1 Mathematical pendulum

6.2.1.1 Double pendulum

Since the planar pendulum example already appeared several times throughout this thesis it does not require any further explanation. The mechanical system shown in Figure 6.1 is a simple extension where a second rod and unit mass have been attached to the first mass point such that there are now two constraints

$$\boldsymbol{g}(\boldsymbol{q}(t)) = \begin{pmatrix} (q_{x,1}(t))^2 + (q_{y,1}(t))^2 - l_1^2 \\ (q_{x,2}(t) - q_{x,1}(t))^2 + (q_{y,2}(t) - q_{y,1}(t))^2 - l_2^2 \end{pmatrix}, \quad l_1 = l_2 = 1\,, \tag{6.4}$$

where we expressed the rigid connections by quadratic relations. In a way, the problem therefore is not 'as strongly nonlinear' as the simple pendulum example and the derived terms as the constraint Jacobian take a much easier form. The double pendulum is a classical example for chaotic systems as may be guessed from the right plot in Figure 6.1 where we depict the time evolution over the time interval $[t_0, t_{\text{end}}] = [0, 5]$. To get rather strong excitations and keep consistent with the simple pendulum we used the initial values $\boldsymbol{q}(t_0 = 0) = \frac{1}{2}(\sqrt{2}, \sqrt{2}, 2\sqrt{2}, 2\sqrt{2})^\top$, $\dot{\boldsymbol{q}}(t_0) = (-1, 1, -1, 1)^\top$.

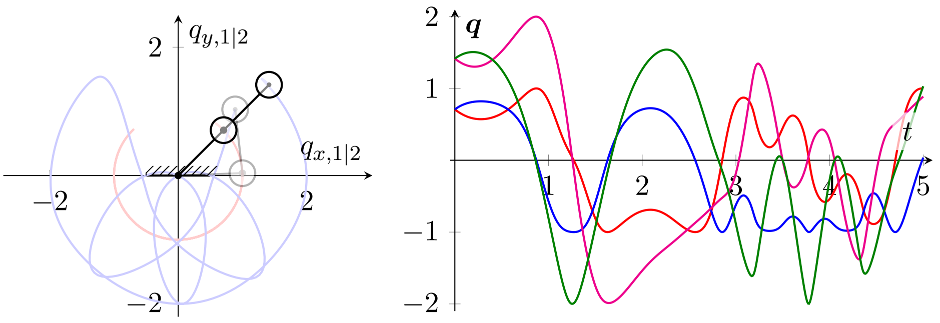


Figure 6.1: Double pendulum: Initial/end configuration and time evolution

6.2.2 Seven body mechanism

“Andrews’ squeezer mechanism” or the seven body mechanism depicted in Figure 6.2 is one of the most well-documented and well-tried realistic benchmark problem in the multibody system community. Schiehlen (1990) used it as the reference for planar constrained mechanical systems in technical simulation and in the monograph of Hairer and Wanner (2002) where the parameters have been adapted from it serves as ideal illustration of a mechanical system. As the name suggests the benchmark describes the (plane and frictionless) motion of seven bodies interconnected to each other by joints that realize three kinematic loops. Initially at rest, a constant torque raises an ever faster motion of the system. In the numerical experiments below we will use two different formulations to describe the motion of the system. The first version is documented in great detail in (Hairer and Wanner, 2002, Lionen et al., 1996) and has $n_{\boldsymbol{q}} = 7$ position variables that identify the angles of the bodies in Figure 6.2 and are constrained by $n_{\boldsymbol{\lambda}} = 6$ nonlinear equations to close the three kinematic loops. These three closed loops, while there are only seven bodies in the system, make Andrews’ squeezer mechanism a highly coupled problem with complicated mass matrix, constraint Jacobian, and force vector. Hence, the second model uses x and y coordinates of the centers of mass and angles of each body as its $n_{\boldsymbol{q}} = 21$ variables. Consequently, there are $n_{\boldsymbol{\lambda}} = 20$ constraint equations necessary to restrain the motion to one single degree of freedom. With altogether 41 state and Lagrange multiplier variables the benchmark may, judging from the dimension, be reckoned a quite realistic problem. The time horizon of the seven body mechanism is usually taken as $[t_0, t_{\text{end}}] = [0, 0.03]$. In Chapter 5 we saw that the convergence behavior of Newmark integrators may depend on the initial values. So, for the ‘large’ seven body mechanism we used a very accurately acquired reference solution to obtain initial values for $t_0 := 0.01$, that were afterwards projected to fulfill position and velocity constraints to machine precision and have non-vanishing initial velocities.

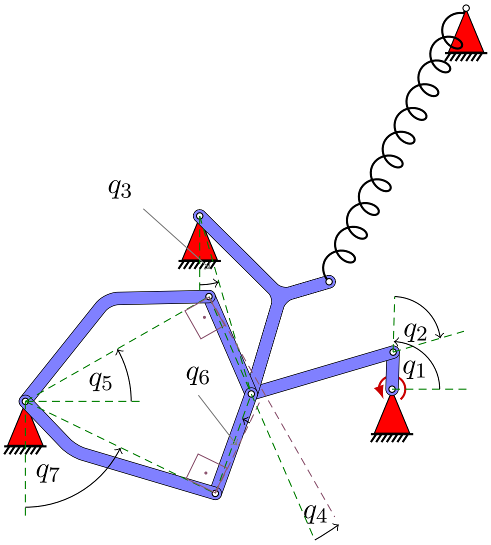


Figure 6.2: Schematic illustration of the seven body benchmark, cf. (Hairer and Wanner, 2002, Schiehlen, 1990)

6.2.3 Slider crank mechanism

This benchmark was proposed by Simeon (1996) to study typical effects of multibody systems in which flexible and rigid structures interact with each other and has become part of an accepted standard benchmark set (Lionen et al., 1996) for ODE and DAE integration methods. In the model a flexible rod moves between a rigid block whose motion is constrained to a fixed line and a rigid rod that rotates around a fixed point with constant angular velocity, see Figure 6.3.

The flexible motion of the rod is described by a Galerkin approach with two ansatz functions for longitudinal and lateral displacement, each one possessing two degrees of freedom and a nonlinear material law, cf. Example 3.19. With the angles of the rods and the position of the block along the line we have $n_{\boldsymbol{q}} = 7$ position coordinates and three constraints: two for the closed loop and one rheonomic constraint for the constant revolution of the rigid rod. The highly oscillatory motion and nonlinear potential model make it an excellent example for the application of methods from structural dynamics.

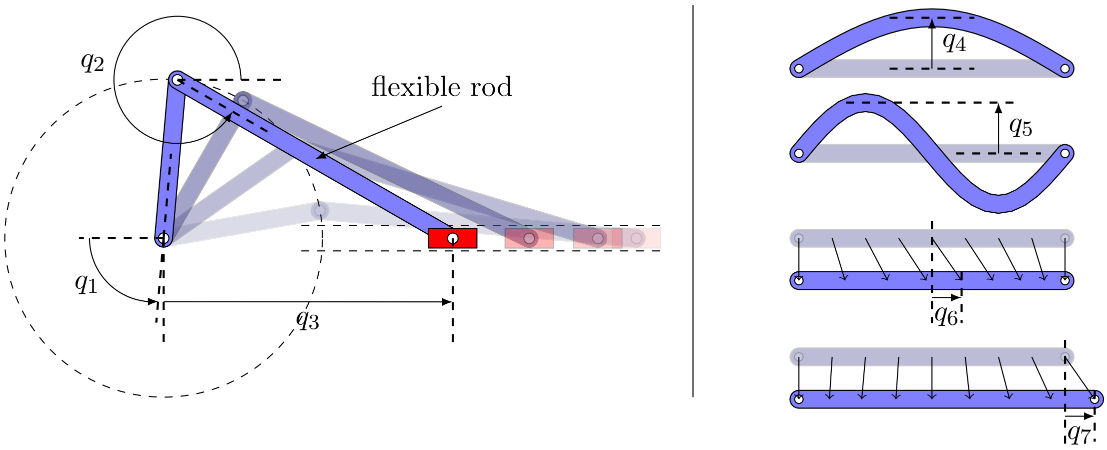


Figure 6.3: Illustration of the slider crank benchmark, Galerkin ansatz functions of the flexible rod, cf. (Simeon (1996), Kettmann (2009))

6.3 DAE systems

Prior to the convergence proofs in the index-2 and index-3 DAE case we have already presented some numerical experiments confirming the convergence results of Corollaries 5.14 and 5.22. So, in this section we will just briefly underline two issues for more realistic benchmark problems before we turn our attention to the SPP case in Sections 6.4 and 6.5 below. For a numerical verification of the convergence results in the DAE case we refer to the literature (e. g. Lunk and Simeon (2006), Arnold and Brüls (2007), Jay and Negrut (2007), Arnold et al. (2016)).

6.3.1 Numerical dissipation

The generalized-α method has been originally introduced by Chung and Hulbert (1993) as an ‘algorithm for structural dynamics with improved numerical dissipation’. The first experiment will address this matter in the DAE setting. According to Cardona and Géradin (1989), typical values of ϱ_∞ for applications in multibody dynamics and the HHT or WBZ method are in the range of $0.55 - 1$. Other researchers use maximum damping, i. e., $\varrho_\infty = 0$ to obtain a maximally stable time integration first and then increase ϱ_∞, mostly for accuracy reasons.

In Figure 6.4 the effect of algorithmically controllable damping is illustrated for the slider crank benchmark problem in its index-3 formulation. Clearly, as shown by Lunk and Simeon

(2006), the beneficial user-definable damping properties carry over from the ODE to the DAE case.

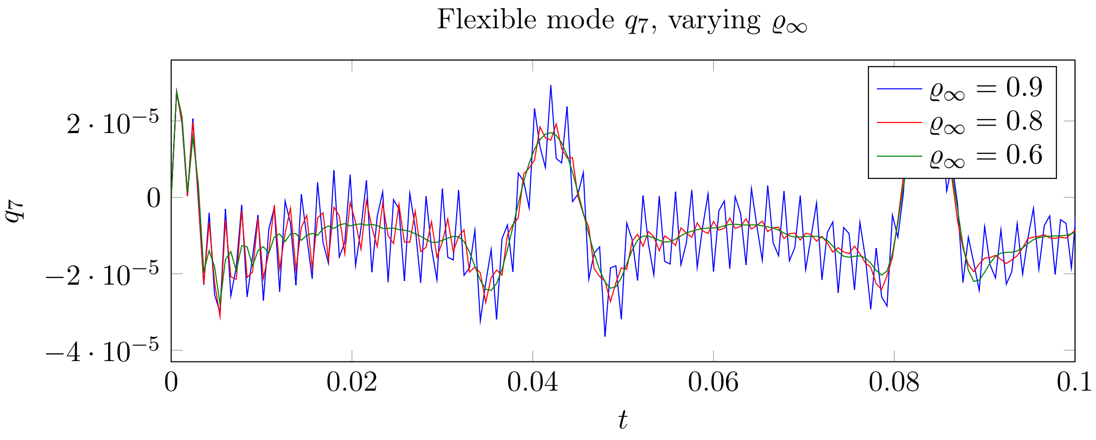


Figure 6.4: Oscillatory mode q_7 of slider-crank mechanism for different choices of numerical damping ϱ_∞ and time step size $h = 6 \cdot 10^{-4}$

6.3.2 Overshoot

In Remark 4.21 we showed that the asymptotic linear error behavior of Newmark-type integrators depends on the Jordan structure of the amplification matrix for the linear oscillator in the limit of infinite stiffness. The classical parameter set of the CH(ϱ_∞) algorithm, despite its clearly beneficial behavior in terms of dissipation and dispersion, is prone to locally quadratic overshoot while HHT and WBZ have only linear asymptotic growth of matrix powers. We also presented an alternative way of choosing the algorithm's parameters such that even this linear growth is no longer present, denoted as Gen(ϱ_∞, ϕ_0).

The upshot of this parameter set with improved behavior in the transient phase is illustrated in Figure 6.5. For the 'large', i. e., $n_{\boldsymbol{q}} = 21$, version of the seven body mechanism in its index-3 formulation we performed three time integrations with time step size $h = 2 \cdot 10^{-5}$ using CH(ϱ_∞), HHT and Gen($\varrho_\infty, 0$). The numerical damping parameter was set to the relatively large value $\varrho_\infty = 0.95$ corresponding to $\alpha_{\text{HHT}} = -\frac{1}{39}$. The experiment shows that overshoot in the Lagrange multipliers can be observed for all three parameter sets, yet the reduction of spurious oscillations for HHT and especially Gen(ϱ_∞, ϕ_0) with $\phi_0 = 0$ is also recognizable. This observation corresponds directly to the amplification behavior depicted in Figure 4.7.

6.4 Strongly damped systems

To verify the results from Section 5.3.1 we use the 'classical' form of the seven body mechanism. Corollary 5.33 gives asymptotic estimates for the global errors on position and velocity level that include the penalty parameter δ and the time step size h. Both dependencies are illustrated in Figure 6.6. The maximum error for $(\boldsymbol{q}, \boldsymbol{v})^\top$ on the time horizon $[0, 0.03]$ is shown vs. h and δ in double logarithmic scale. The left plot affirms the second order convergence of the method. For very small time step sizes the error saturates at a level that grows linearly with the perturbation parameter δ. Conversely, in the right plot the first order convergence in δ is apparent and the error saturates at a level in $\mathcal{O}(h^2)$. For this experiment we chose consistent initial values $\boldsymbol{q}_0^\delta$, $\boldsymbol{v}_0^\delta$.

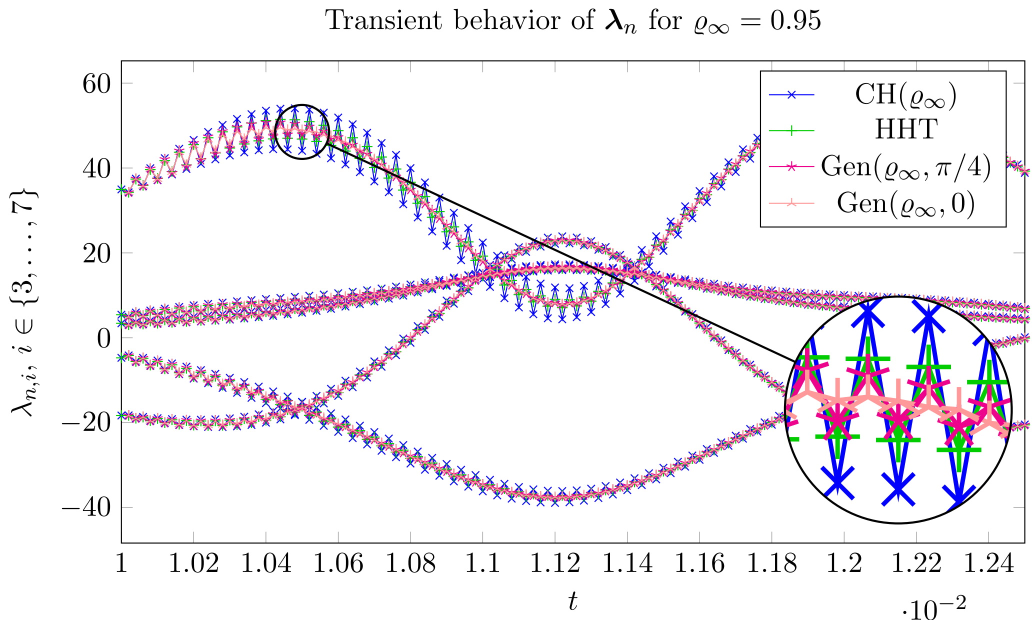


Figure 6.5: Improved transient behavior for amplification matrices with different Jordan structures: Lagrange multipliers of the (large) seven body mechanism

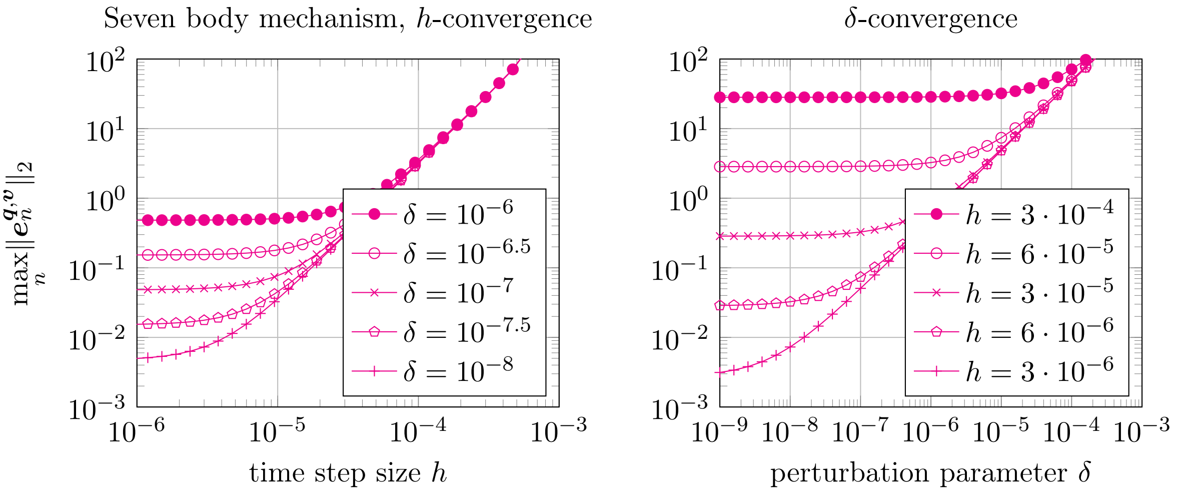


Figure 6.6: Verification of error bounds from Corollary 5.33

To lay out the possible modifications we perform the same experiment a second time with the following alterations:

- For the experiment in Figure 6.6 we used the parameter set CH(0.75). Now, the algorithmic parameters are chosen as $\alpha_m = \frac{2}{7}$, $\alpha_f = \frac{3}{7}$, $\gamma = \frac{9}{14}$, $\beta = \frac{1}{7}$. The method is zero stable and second order accurate but lacks all other stability concepts from Definition 4.7. For sufficiently large stiffness the algorithm would even fail to integrate the harmonic oscillator in a stable way.
- We also perturbed the initial values according to the limits given by Assumption 5.25 and added perturbation terms of size $\mathcal{O}(h^2)$ and $\mathcal{O}(\delta)$ on position and velocity level, respectively.

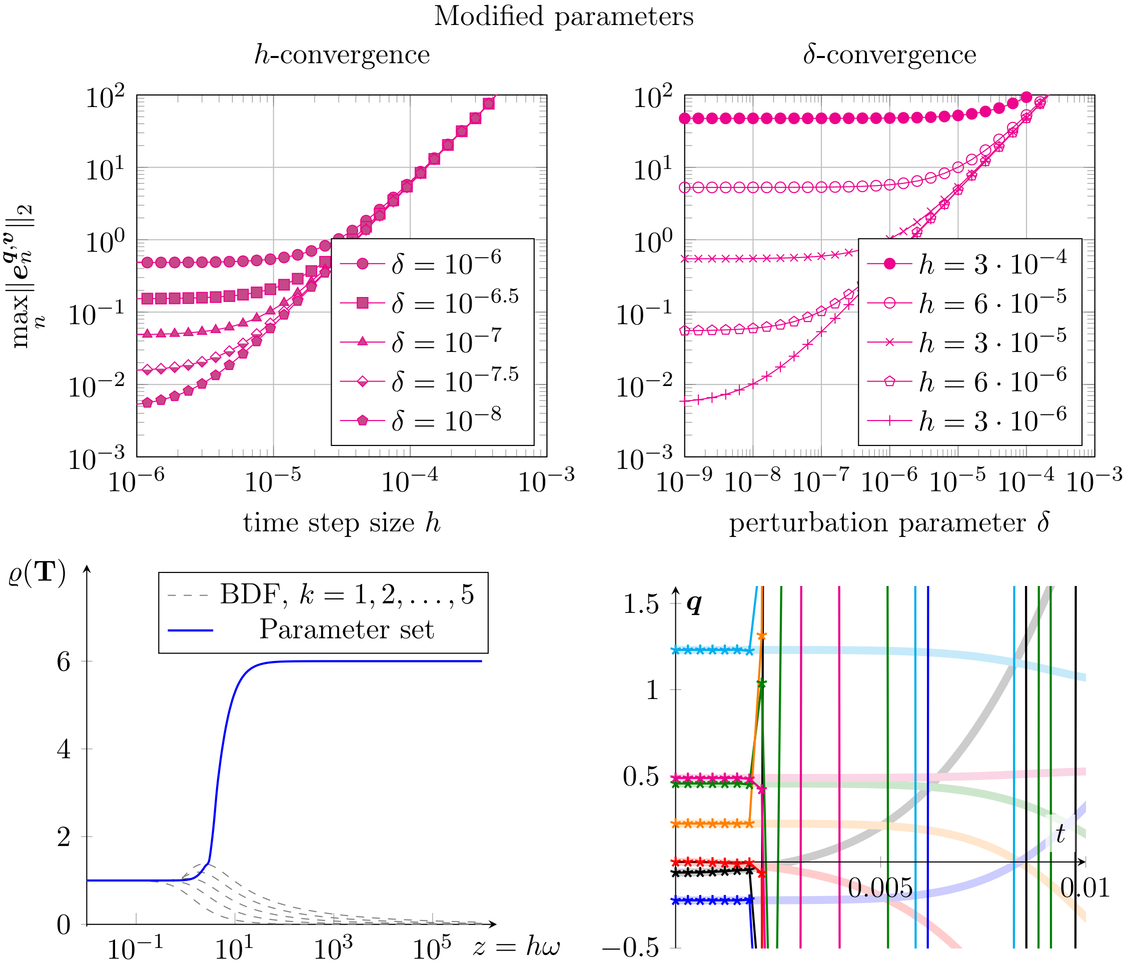


Figure 6.7: Damped squeezer mechanism: step size convergence for perturbed initial values and the parameter set $\alpha_m = \frac{2}{7}$, $\alpha_f = \frac{3}{7}$, $\gamma = \frac{9}{14}$, $\beta = \frac{1}{7}$, second row: amplification factor of the algorithm for the harmonic oscillator and failed integration of the corresponding stiff mechanical system

In the lower left plot of Figure 6.7 we illustrate the numerical damping properties of the algorithm used, cf. the corresponding plots in Figures 4.1, 4.3, 4.6 and 4.9. It is ‘unconditionally unstable’ (Vater et al., 2011) as for no step size $h > 0$ the error amplification in the linear regime is contractive. Nevertheless, the numerical results in the upper plots match the behavior in the prior example since the 'unusual' parameter set meets the requirements of Theorem 5.32. To underline this result we used the same algorithmic parameters for the stiff formulation with the moderate value $\varepsilon = 10^{-3}$ and time step size $h = 3 \cdot 10^{-4}$. The result is shown in the lower right plot of Figure 6.7. The light colors indicate a reference solution of the stiff system for $\varepsilon = 10^{-5}$ and the parameter set of Chung and Hulbert (1993). Clearly, the algorithm completely fails to give a stable solution in the stiff case.

6.5 Stiff systems

The theoretical findings of Section 5.3.2 are underlined by means of the simple/double pendulum example. Figure 6.8 verifies the second order of convergence for position variables by means of time step size h and penalty parameter ε. The parameter set CH(0.8) is used in this case.

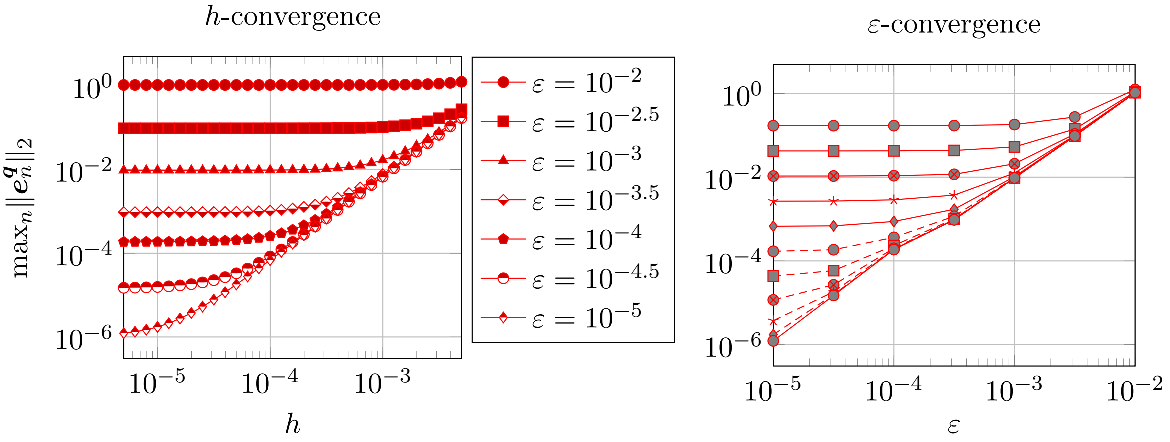
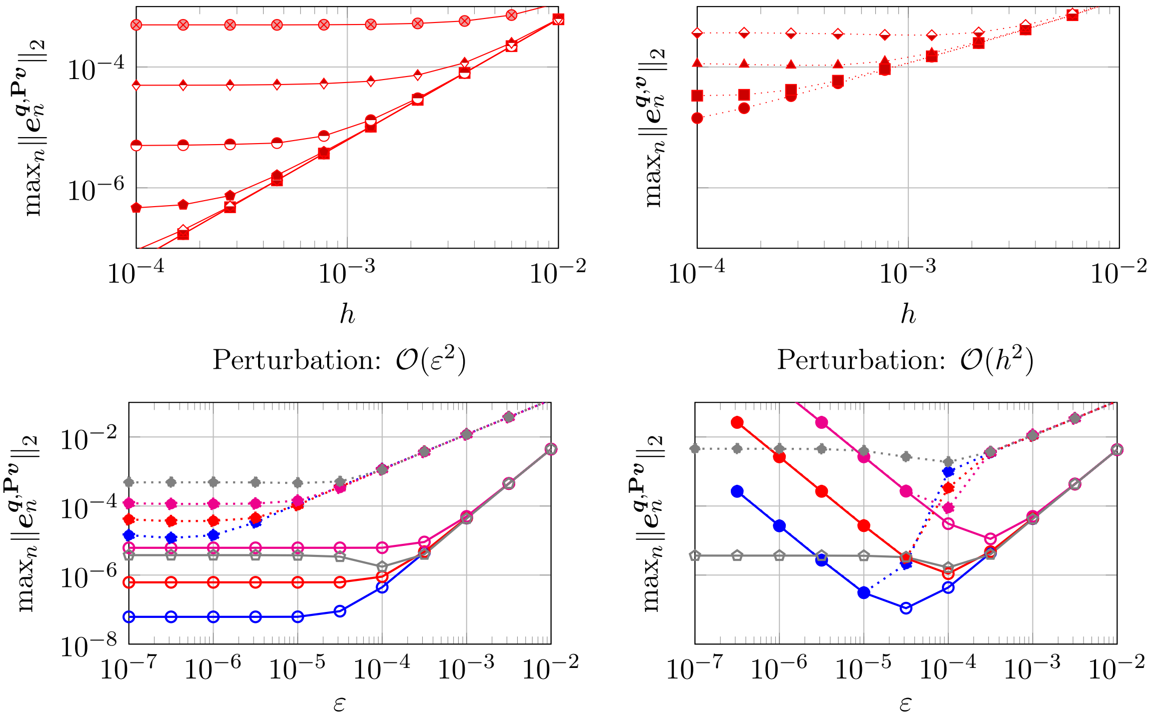


Figure 6.8: Numerical convergence behavior for the double pendulum example as stiff mechanical system

To verify the order reduction for the orthogonal velocity error components we come back to the simple pendulum benchmark and display the maximum error over the time interval $[0, 2]$ in Figure 6.9. We also changed the initial conditions to $\boldsymbol{q}(t_0) = (0, -1)^\top$, $\boldsymbol{v}(t_0) = (1, 0)^\top$. In the first row the convergence for fixed ε and $h \to 0$ is displayed. Omitting the orthogonal error component $\boldsymbol{e}_n^{\mathbf{G}\boldsymbol{v}}$ from the condensed error term $\boldsymbol{e}_n^{\boldsymbol{q},\boldsymbol{v}} := (\boldsymbol{e}_n^{\boldsymbol{q}}, \boldsymbol{e}_n^{\boldsymbol{v}})^\top$ 'increases' the order of convergence of the algorithm from one to two. For the results in Figure 6.9 we used a numerical damping of $\varrho_\infty = 0.9$. Initial values on position level were disturbed by $\mathcal{O}(\varepsilon^2)$. The different plots indicate different penalty parameters ε ranging from 10^{-2} to 10^{-7}. As expected, the error saturates if the step sizes become very small.

In the first plot of the second row we display the accuracy of the numerical approximations as a function of ε (the same experiment). Dashed plots indicate $\boldsymbol{e}_n^{\boldsymbol{q},\boldsymbol{v}}$ while solid lines show the error behavior of the projected terms $\boldsymbol{e}_n^{\boldsymbol{q},\mathbf{P}\boldsymbol{v}}$. Qualitatively we get the same results as for the previous experiment, i. e., second order of convergence with respect to ε and saturation at a level in $\mathcal{O}(h^2)$. To show again that the requirements on the initial values in Assumption 5.25 are sharp we show how convergence can no longer be seen if the deviation from $\boldsymbol{g} = \mathbf{0}$ is growing quadratically with the time step size, see the lower right plot. The gray lines in both plots indicate the according results for the SDIRK method we constructed in Remark 4.22. As the results of Lubich (1993) from Remark 3.20 indicate, even a deviation from the constraint manifold by $\mathcal{O}(h^2)$ does not



Perturbation: $\mathcal{O}(\varepsilon^2)$

Perturbation: $\mathcal{O}(h^2)$

Figure 6.9: Verification of error bounds for the pendulum example

impair the convergence. Yet, having a look at (3.26) again, we see that the order of the Runge–Kutta method drops to one since it is only of stage order $p_\mathrm{s} = 1$. One should also keep in mind that the computational effort of this algorithm is roughly twice as large since one has to solve two nonlinear systems in each time step to obtain the stage vectors.

Constraint fulfillment and drift At last, to illustrate that strongly damped systems are not superior from all perspectives of numerical analysis to stiff mechanical systems (no order reduction occurs) we come back to the double pendulum benchmark problem. Figure 6.10 shows the norm of the residual in the constraint equation (6.4) as it evolves throughout the time integration. To have a larger effect in the transient phase the initial values were perturbed by $\mathcal{O}(h^2)$ in $\boldsymbol{g}(\boldsymbol{q}_0^{\varepsilon|\delta})$ and $\mathcal{O}(h)$ in its time derivative $\mathbf{G}(\boldsymbol{q}_0^{\varepsilon|\delta})\boldsymbol{v}_0^{\varepsilon|\delta}$. For the stiff model we used $\varepsilon = 10^{-4}$, for the damped system a penalty parameter $\delta = 10^{-6}$ was chosen with a time step size $h = 0.005$ in both cases. In particular, in the plots three phenomena become apparent: (a) Again we see that smaller values of ϱ_∞ impose a stronger numerical damping, in the constraint evolution and also for the strongly damped case, (b) in neither of the formulations do the constraint residuals tend exactly towards zero. They seem to approach the smooth motion introduced in Theorems 3.8 and 3.16 instead as it is the case for Runge–Kutta methods as well, and (c) the drift-off phenomenon known from index-reduced DAE systems transfers to the singularly perturbed counterparts as well. Note how *the imposed weak constraint* ($\mathbf{G}\dot{\boldsymbol{q}}$ for strongly damped, $\boldsymbol{g}$ for stiff systems) remains within a neighborhood of zero whose size depends on the penalty parameter. The high similarity of the first and the last plot is fairly coincidental and due to the solution progression with its sharp peaks, cf. Figure 6.1.

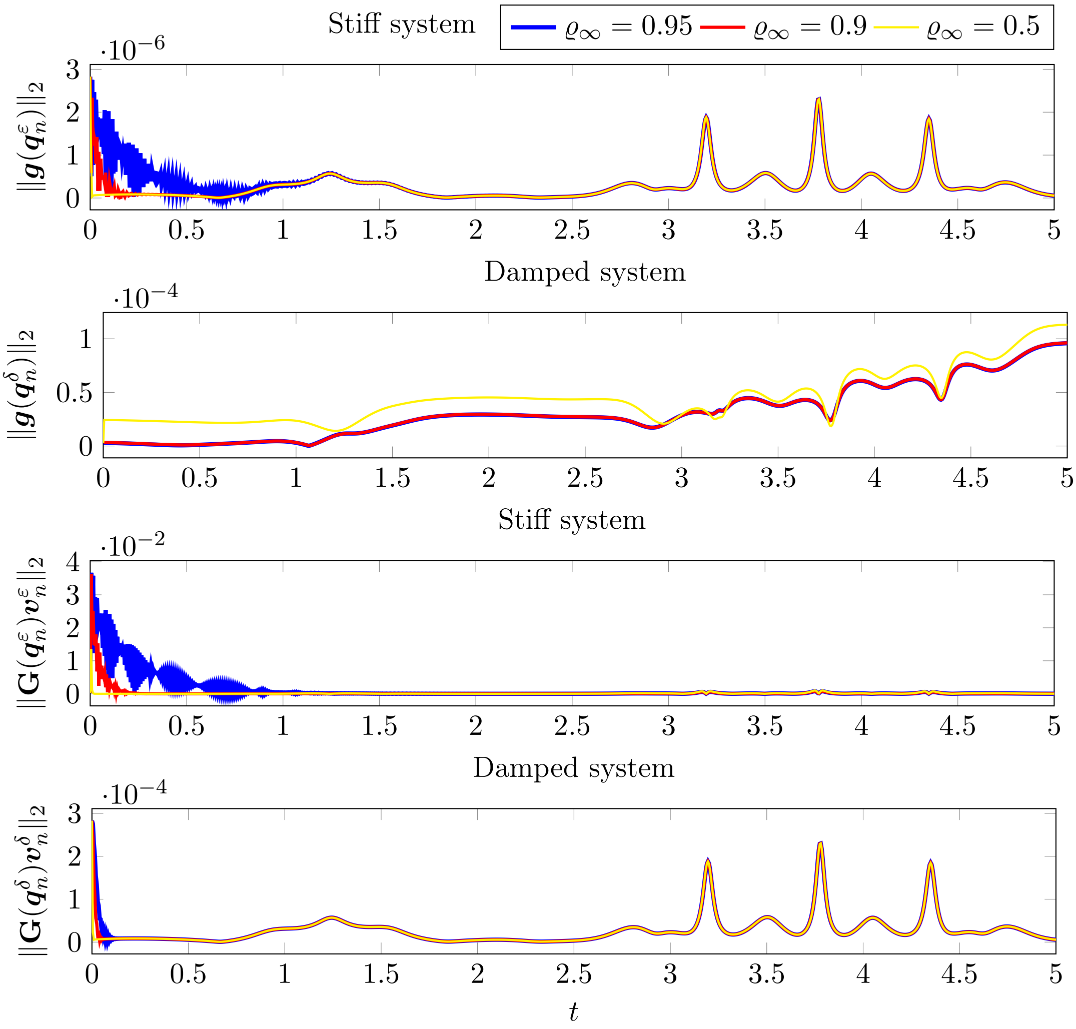


Figure 6.10: Time evolution of position and velocity constraints for the double pendulum in strongly damped and stiff formulation for varying numerical damping

6.6 Sidetrip: Remedies in case of large initial deviations: The HMM approach

Compared to the convergence results of Lubich (1993) and Stumpp (2006) for Runge–Kutta-type methods the conditions on the initial values in Assumption 5.25 seem rather restrictive. Apart from the transient phase the Newmark-type integrators outperform common Runge–Kutta algorithms and keep their second order of convergence while also enabling the user to adapt the numerical damping. This section shall collect a few ideas on how to resolve the problem that rather strong conditions on the initial values are needed for convergence. Nevertheless, it shall merely serve as a lookout on aspects of further investigations. In lack of Assumption 5.25 the error analysis of the last chapter looses its validity. For simplicity we will restrict the experiments in this section solely to stiff mechanical systems which is the more important and challenging case anyway.

Lemma 5.39 shows that the most important problem lays in the bad approximation from the first time step on. If the manifold of the slow motion is known it is, of course, possible to project the initial values $(\boldsymbol{q}_0^\varepsilon, \boldsymbol{v}_0^\varepsilon)^\top \rightsquigarrow (\bar{\boldsymbol{q}}_0, \bar{\boldsymbol{v}}_0)^\top := (\boldsymbol{\pi}_0(\boldsymbol{q}_0^\varepsilon), \mathbf{P}_0\boldsymbol{v}_0^\varepsilon)^\top$ onto $T\,\mathfrak{M}^{\mathrm{s}}$ and the problems for large initial deviations are shed. But what is there to do if a direct projection is not known?

In Chapter 2 we have introduced the concept of heterogeneous multiscale methods (HMM). Following the general framework, the stiff mechanical system is the *microsystem* which has to be integrated on a small time interval $[t_0, t_0{+}\eta]$, $(\eta > 0)$, (e. g. with the same method but very small time steps) and the acquired data can then be used to approximate the *macrosystem*, i. e., the slow or smooth motion of the mechanical system, cf. Figure 6.11. For the projection procedure

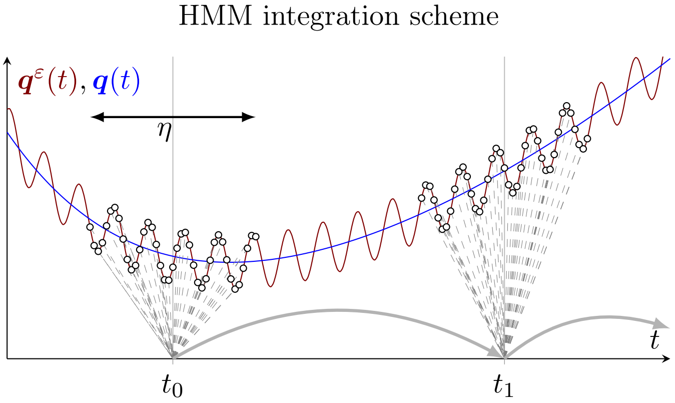


Figure 6.11: The HMM approach for highly oscillatory integrands (Abdulle et al., 2012)

we follow the work of Ariel et al. (2012). Since the analytic solutions of (3.15) oscillate around the slow motion a simple average

$$\bar{\boldsymbol{q}} := \int_{t_0}^{t_0+\eta} \boldsymbol{q}^\varepsilon(t; \boldsymbol{q}_0^\varepsilon, \boldsymbol{v}_0^\varepsilon)\ \mathrm{d}t \quad \bar{\boldsymbol{v}} := \int_{t_0}^{t_0+\eta} \dot{\boldsymbol{q}}^\varepsilon(t; \boldsymbol{q}_0^\varepsilon, \boldsymbol{v}_0^\varepsilon)\ \mathrm{d}t\,,$$

where the notation $\boldsymbol{q}^\varepsilon(t;\cdot,\cdot)$ indicates the dependency on the initial values, usually suffices to obtain smoother solutions. A more detailed analysis (E and Engquist, 2003) shows that the approximation can be improved if instead of a simple average, a kernel function $\mathcal{K}_\eta(\cdot)$ is multiplied with the above integrand and the integration interval is symmetric around t_0. Using this technique for the pendulum example from Chapter 1 gives similar results as the ones for initial perturbations in $\mathcal{O}(\varepsilon^2)$ we have already seen.

In Remarks 5.17 and 5.23 we showed that adapted initial values for the acceleration-like variables $\boldsymbol{a}_0$ may prevent order reduction in the transient phase for the DAE case. Using difference approximations for the third derivative like

$$\boldsymbol{a}_0 := \ddot{\boldsymbol{q}}(t_0) + h\Delta_\alpha \frac{\ddot{\boldsymbol{q}}(t_0 + \eta/2) - \ddot{\boldsymbol{q}}(t_0 - \eta/2)}{\eta}\,,$$

for some given constant $\eta > 0$ resolved the problem of order reduction in the Lagrange multipliers in the index-2 case and was one ingredient to overcome the drop to order one for the velocity variables in index-3 systems. So, the above procedure might as well be used to acquire this kind of correction term, i. e., using the approximants

$$\ddot{\boldsymbol{q}}(t_0 \pm \eta/2) \approx \mp \int_{t_0\pm\eta}^{t_0} [\mathcal{K}_\eta \cdot \mathrm{rhs}](t_0 \pm \tfrac{\eta}{2} + t)\ \mathrm{d}t\,, \quad \ddot{\boldsymbol{q}}(t_0) \approx \int_{t_0-\eta/2}^{t_0+\eta/2} [\mathcal{K}_\eta \cdot \mathrm{rhs}](t_0 + t)\ \mathrm{d}t\,, \qquad (6.5)$$

with $\mathrm{rhs}(t) := \mathbf{M}^{-1}(\boldsymbol{f} - \frac{1}{\varepsilon^2}\mathbf{G}^\top \boldsymbol{g})$.

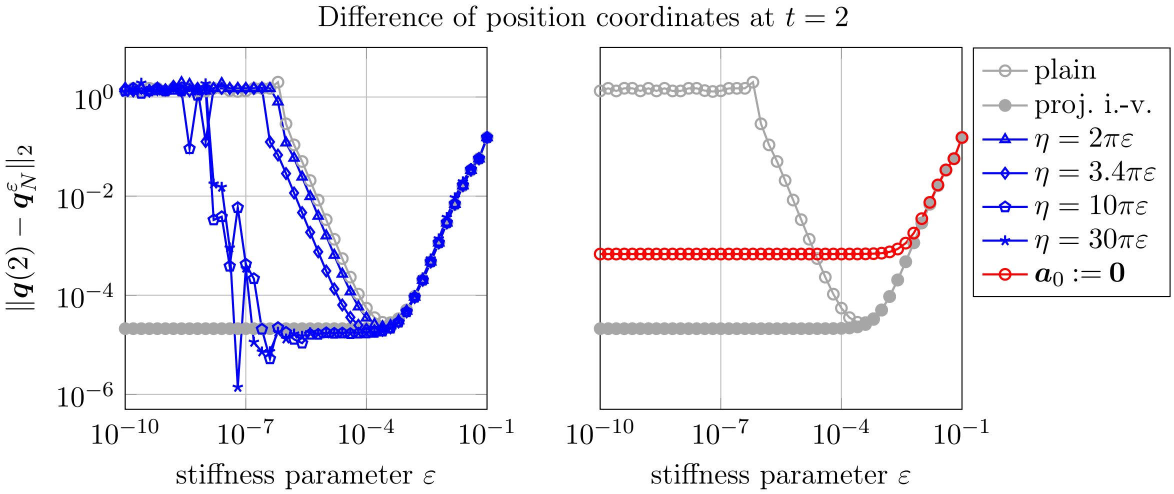


Figure 6.12: Improved convergence for the pendulum example using HMM (left) or zero initial-acceleration (right), solutions for classical initialization (plain) or prior projection of initial values (proj. i.-v.)

In Figure 6.12 we depict the result for the above procedure using the common kernel function $\mathcal{K}_\eta(t) := \frac{1}{\eta}(1 + \cos(2\pi(t - t_0)/\eta))$. The gray plots are the results from Chapter 1 for $\mathcal{O}(h^2)$ perturbation and projected initial values. Micro integrations were performed using a standard integration method with very low tolerances and a trapezoidal rule to approximate the integrals in (6.5). It is evident that the adaptation of the initial values on acceleration level may lower the negative influence of the initial perturbations. Even the very rough approximation $\boldsymbol{a}_0 := \boldsymbol{0}$ which is drawn in the right plot of Figure 6.12 keeps the numerical results from diverging at the cost of order reduction in $\boldsymbol{q}$. These rather simple considerations show that there are many possibilities to further improve the performance of the Newmark family algorithms in the SPP setting.

Chapter 7

Summary

Singularly perturbed problems represent some of the most important challenges in the numerical simulation of technical systems. In this thesis we analyzed the widely used family of Newmark-type methods including the classical Newmark-scheme, HHT and the CH(ϱ_∞)-generalized-α method for the two important classes of strongly damped and stiff mechanical systems.

To this means, we underlined the interrelation between the limit cases of constrained mechanical systems in DAE form and the linear special cases known from classical stability analysis of time integration methods in technical simulation. We saw in particular that order reduction in the Lagrange multipliers in the index-2 case can be avoided by an appropriate initialization of the acceleration-like variables $\boldsymbol{a}_0$ while in the index-3 case a real adaptation is necessary. These drawbacks carry over to the SPP setting where in the strongly damped case the order reduction result does not influence the convergence behavior on the relevant position and velocity level but for stiff mechanical systems remains apparent and causes a transient reduction to first order for the velocities. Finding remedies in the singularly perturbed case appears to be way more challenging since initial errors in $\mathcal{O}(1)$ are involved. Even so, a comprehensive error analysis that based on a onestep representation of the algorithm has been carried out and convergence in time step size as well as penalty parameters and bounds for the initial deviations have been given.

One of the main tools in the error analysis was the close relationship to the DAE case and, accordingly, the local truncation as well as global errors were defined with respect to the slow motion of the mechanical systems whose existence is guaranteed from the Rubin–Ungar Theorem. Secondly, it proved beneficial to rely on the onestep representation of the algorithms to derive recursion formulae such that in all four cases the synthesis of the error analysis could have been carried out within exactly the same framework. The governing amplification behavior coincides with the one from the scalar-valued linear regime which we analyzed and visualized in great detail providing a rather comprehensive overview on the theoretical and practical developments concerning this family of algorithms in the last decades.

The main differences of constrained and singularly perturbed problems can be classified by the large initial errors and the additional coupling terms of artificial Lagrange multipliers and acceleration-like variables in the strongly damped case and even the velocity components for stiff mechanical systems. The latter ones lead to the necessity of dealing with amplification matrices that depend on the ratios of penalty parameters and time step sizes. As a result the guaranteed contractivity of the error recursion implies an upper bound on those relations.

The introduction of artificial Lagrange multipliers not only proves very useful for theoretical purposes but is also the method of choice if one is to practically implement the algorithms since the singular force terms forbid a reliable corrector iteration that is solely based on the ODE formulation of the problems. In Chapter 6 we presented a solution procedure by means of a pseudocode and demonstrated the validity of the theoretical finding by a series of numerical experiments of rather small dimension. They were, nevertheless, sufficient to show that the bounds in the error estimates are sharp and if any of the required assumption falls the algorithms may no longer converge at all.

The numerical experiments, nonetheless, indicate that the seemingly divergent behavior of Newmark-type integrators is no unstable behavior in the sense that the numerical approximations grow large beyond any bounds or that the solution of the nonlinear systems fails. It is rather to expect that further research on the topic might reveal a similar analytic background as can be shown for onestep methods and that the smooth motion of the mechanical system is more appropriate to measure the performance of the methods as opposed to the slow one. From a practical point of view there are also a few other topics that have not been discussed so far. The debatably most important concern is the analysis and practical test of variable time step integration methods in the setting of this thesis but also possible benefits from using mixed formulations, cf. Remark 3.21, or the consideration of nonlinear configuration spaces or the connection to questions of structure-preservation have found much interest in recent years.

After all, the application of Newmark-type integrators to real-world problems within the SPP framework of this thesis will show to which extend the results we have presented prove useful.

Bibliography

A. Abdulle. Fourth order Chebyshev methods with recurrence relation. *SIAM J. Sci. Comput.*, 23(6):2041–2054, 2002. doi: 10.1137/S1064827500379549. One citation on page 25.

A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden. Heterogeneous multiscale methods. *Acta Numerica*, 21:1–87, 2012. doi: 10.1017/S0962492912000025. 2 citations on pages 25 and 125.

R. Altmann. *Regularization and simulation of constrained partial differential equations*. PhD thesis, Technical University of Berlin, 2015. `https://opus4.kobv.de/opus4-tuberlin/frontdoor/index/index/docId/6699`. 2 citations on pages 27 and 108.

G. Ariel, J.M. Sanz-Serna, and R. Tsai. A multiscale technique for finding slow manifolds of stiff mechanical systems. *Multiscale Model. Simul.*, 10(4):1180–1203, 2012. doi: 10.1137/120861461. 3 citations on pages 26, 28, and 125.

L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific J Math.*, 16(1):1–3, 1966. doi: 10.2140/pjm.1966.16.1. One citation on page 115.

M. Arnold. The generalized-α method in industrial multibody system simulation. In K. Arczewski, J. Frączek, and M. Wojtyra, editors, *Proceedings of Multibody Dynamics 2009 (ECCOMAS Thematic Conference), Warsaw, Poland, June 29th - July 2nd*, 2009. 2 citations on pages 77 and 86.

M. Arnold. DAE aspects of multibody dynamics. A. Ilchmann and T. Reis, editors, Differential-Algebraic Equations Forum. Springer, 2016. submitted, a preliminary version of this material was published as Technical Report 01-2016, Martin Luther University Halle–Wittenberg, Institute of Mathematics. One citation on page 16.

M. Arnold and O. Brüls. Convergence of the generalized-α scheme for constrained mechanical systems. *Multibody Syst. Dyn.*, 18:185–202, 2007. doi: 10.1007/s11044-007-9084-0. 9 citations on pages 49, 55, 57, 73, 75, 110, 112, 114, and 118.

M. Arnold and St. Hante. Implementation details of a generalized-α DAE Lie group method. *J. Comp. Nonlin. Dyn.*, 2016. doi: 10.1115/1.4033441. in print. 2 citations on pages 86 and 110.

M. Arnold, B. Burgermeister, and A. Eichberger. Linearly implicit time integration methods in real-time applications: DAEs and stiff ODEs. *Multibody Syst. Dyn.*, 17(2):99–117, 2007. doi: 10.1007/s11044-007-9036-8. One citation on page 114.

M. Arnold, B. Burgermeister, C. Führer, G. Hippmann, and G. Rill. Numerical methods in vehicle system dynamics: state of the art and current developments. *Vehicle Syst. Dyn.: Int. J. Vehicle Mech. Mobility*, 49(7):1159–1207, 2011. doi: 10.1080/00423114.2011.582953. 6 citations on pages 7, 19, 50, 62, 110, and 113.

M. Arnold, O. Brüls, and A. Cardona. Error analysis of generalized-α Lie group time integration methods for constrained mechanical systems. *Numer. Math.*, 129:149–179, 2015a. doi: 10.1007/s00211-014-0633-1. 10 citations on pages 72, 73, 74, 78, 79, 84, 87, 88, 89, and 103.

M. Arnold, O. Brüls, and A. Cardona. Order reduction in time integration caused by velocity projection. *J. Mech. Sci. Tech.*, 29(7):2579–2585, 2015b. doi: 10.1007/s12206-015-0501-7. revised and extended version published as Technical Report 02-2015, Martin Luther University Halle-Wittenberg, Institute of Mathematics. One citation on page 88.

M. Arnold, A. Cardona, and O. Brüls. A Lie algebra approach to Lie group time integration of constrained systems. In P. Betsch, editor, *Structure-Preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics*, volume 565 of *CISM Courses and Lectures*, pages 91–158. Springer International Publishing, 2016. doi: 10.1007/978-3-319-31879-0_3. Preliminary version, published as Technical Report 01-2015, Martin Luther University Halle-Wittenberg, Institute of Mathematics. 8 citations on pages 53, 73, 85, 86, 91, 113, 114, and 118.

V.I. Arnold. *Geometrical Methods in the theory of ordinary differential equations.* Number 250 in Grundlehren der mathematischen Wissenschaften. Springer New York, 2nd edition, 1988. doi: 10.1007/978-1-4612-1037-5. 2 citations on pages 9 and 26.

V.I. Arnold. *Mathematical methods of classical mechanics.* Number 60 in Graduate Texts in Mathematics. Springer New York, 2nd edition, 1989. 2 citations on pages 37 and 38.

U.M. Ascher, H. Chin, and S. Reich. Stabilization of DAEs and invariant manifolds. *Numer. Math.*, 67(2):131–149, 1994. doi: 10.1007/s002110050020. One citation on page 18.

U.M. Ascher, H. Huang, and K. van den Doel. Artificial time integration. *BIT Numer. Math.*, 47(1):3–25, 2007. doi: 10.1007/s10543-006-0112-x. One citation on page 116.

A. Barrlund. Constrained least squares methods for linear time varying DAE systems. *Numer. Math.*, 60(1):145–161, 1991. doi: 10.1007/BF01385719. One citation on page 18.

K.J. Bathe and E.L. Wilson. Stability and accuracy analysis of direct integration methods. *Earthq. Engrg. Struct. Dyn.*, 1(3):283–291, 1973. doi: 10.1002_eqe.4290010308. One citation on page 63.

J. Baumgarte. Stabilization of constraints and integrals of motion in dynamical systems. *Comp. Meth. Appl. Mech. Engrg.*, 1:1–16, 1972. doi: 10.1016/0045-7825(72)90018-7. One citation on page 18.

E. Bayo, J. Garcia de Jalon, and M.A. Serna. A modified Lagrangian formulation for the dynamic analysis of constrained mechanical systems. *Comp. Meth. Appl. Mech. Engrg.*, 71:183–195, 1988. doi: 10.1016/0045-7825(88)90085-0. 2 citations on pages 37 and 44.

G. Bazzi and E. Anderheggen. The ρ-family of algorithms for time-step integration with improved numerical dissipation. *Earthq. Engrg. Struct. Dyn.*, 10:537–550, 1982. doi: 10.1002/eqe.4290100404. 2 citations on pages 62 and 71.

U. Becker. *Efficient time integration and nonlinear model reduction for incompressible hyperelastic materials.* PhD thesis, Technical University of Kaiserslautern, 2012. 2 citations on pages 43 and 70.

U. Becker, B. Simeon, and M. Burger. On Rosenbrock methods for the time integration of nearly incompressible materials and their usage for nonlinear model reduction. *J. Comput. Appl. Math.*, 262:333–345, 2014. doi: 10.1016/j.cam.2013.10.042. 3 citations on pages 29, 70, and 102.

F.A. Bornemann. *Homogenization in time of singularly perturbed mechanical systems*, volume 1687 of *Lecture Notes in Mathematics*. Springer Berlin, 1998. doi: 10.1007/BFb0092091. 4 citations on pages 27, 30, 38, and 42.

C.L. Bottasso, D. Dopico, and L. Trainelli. On the optimal scaling of index three DAEs in multibody dynamics. *Multibody Syst. Dyn.*, 19(1):3–20, 2008. doi: 10.1007/s11044-007-9051-9. 2 citations on pages 111 and 114.

K.E. Brenan, S.L. Campbell, and L.R. Petzold. *Numerical Solution of initial-value problems in differential-algebraic equations*. SIAM, 1996. doi: 10.1137/1.9781611971224. 4 citations on pages 11, 12, 14, and 25.

B. Brogliato. Inertial couplings between unilateral and bilateral holonomic constraints in frictionless Lagrangian systems. *Multibody Syst. Dyn.*, 29(3):289–325, 2013. doi: 10.1007/s11044-012-9317-8. One citation on page 19.

O. Brüls. *Integrated Simulation and Reduced-Order Modeling of Controlled Flexible Multibody Systems*. PhD thesis, University of Liège, Belgium, 2005. 7 citations on pages 16, 50, 51, 52, 63, 77, and 113.

O. Brüls and M. Arnold. The generalized-α scheme as a linear multistep integrator: Towards a general mechatronic simulator. *J. Comput. Nonlinear Dynam.*, 3(4):041007 (10 pages), 2008. doi: 10.1115/1.2960475. 3 citations on pages 50, 61, and 64.

O. Brüls and J.-C. Golinval. The generalized-α method in mechatronic applications. *Z. Angew. Math. Mech.*, 86(10):748–758, 2006. doi: 10.1002/zamm.200610283. One citation on page 50.

B. Brumm and D. Weiss. Heterogeneous multiscale methods for highly oscillatory mechanical systems with solution-dependent frequencies. *IMA J. Numer. Anal.*, 34:55–82, 2014. doi: 10.1093/imanum/drt010. One citation on page 28.

M.P. Calvo and J.M. Sanz-Serna. Instabilities and inaccuracies in the integration of highly oscillatory problems. *SIAM J. Sci. Comput.*, 31(3):1653–1677, 2009. doi: 10.1137/080727658. One citation on page 26.

S.L. Campbell. A general form for solvable linear time varying singular systems of differential equations. *SIAM J. Math. Anal.*, 18(4):1101–1115, 1987. doi: 10.1137/0518081. One citation on page 18.

A. Cardona and M. Géradin. Time integration of the equations of motion in mechanism analysis. *Compututers & Structures*, 33(3):801–820, 1989. doi: 10.1016/0045-7949(89)90255-1. 3 citations on pages 65, 77, and 118.

A. Cardona and M. Géradin. Numerical integration of second order differential-algebraic systems in flexible mechanism dynamics. In M.F.O Seabra Pereira and J.A.C. Ambrósio, editors, *Computer-Aided Analysis of Rigid and Flexible Mechanical Systems*, volume E–268 of *NATO ASI series*, pages 501–529. Kluwer Academic Publishers, Dordrecht, 1994. doi: 10.1007/978-94-011-1166-9_16. 4 citations on pages 25, 37, 63, and 114.

E.A. Celaya and J.J. Anza. BDF-α: A multistep method with numerical damping control. *Univ. J. Comp. Math.*, 1(3):96–108, 2013. doi: 10.13189/ujcmj.2013.010305. One citation on page 70.

J. Chung and G.M. Hulbert. A time integration algorithm for structural dynamics with improved numerical dissipation: the generalized-α-method. *J. Appl. Mech.*, 60:371–375, 1993. doi: 10.1115/1.2900803. 15 citations on pages 2, 24, 48, 49, 52, 60, 64, 66, 82, 94, 100, 108, 113, 118, and 122.

K.D. Clark. A structural form for higher-index semistate equations. I. theory and applications to circuit and control theory. *Linear Algebra and its Appl.*, 98:169–197, 1988. doi: 10.1016/0024-3795(88)90164-4. One citation on page 13.

D. Cohen, T. Jahnke, K. Lorenz, and Chr. Lubich. Numerical integrators for highly oscillatory Hamiltonian systems: A review. In A. Mielke, editor, *Analysis, Modeling and Simulation of Multiscale Problems*, pages 553–576. Springer Berlin Heidelberg, 2006. doi: 10.1007/3-540-35657-6_20. One citation on page 25.

COMSOL Multiphysics version 3.5a, User's Guide. COMSOL Inc., 2008. downloaded from `http://math.nju.edu.cn/help/mathhpc/doc/comsol/guide.pdf` on October 5th 2015. One citation on page 51.

COMSOL Multiphysics version 4.3a, Reference Guide. COMSOL Inc., 2012. downloaded from `http://nf.nci.org.au/facilities/software/COMSOL/4.3a/doc/pdf/mph/COMSOLMultiphysicsReferenceGuide.pdf` on October 5th 2015. One citation on page 51.

P. Console and E. Hairer. Long-term stability of symmetric partitioned linear multistep methods. In L. Dieci and N. Guglielmi, editors, *Current Challenges in Stability Issues for Numerical Differential Equations*, volume 2082 of *Lecture Notes in Mathematics*, pages 1–37. Springer International Publishing, 2014. doi: 10.1007/978-3-319-01300-8_1. One citation on page 24.

C.F. Curtiss and J.O. Hirschfelder. Integration of stiff equations. *Proc. Natl. Acad. Sci. USA*, 38(3):235–243, 1952. doi: 10.1073/pnas.38.3.235. One citation on page 24.

G. Dahlquist. Convergence and stability in the numerical integration of ordinary differential equations. *Math. Scand.*, 4:235–243, 1956. One citation on page 54.

G. Dahlquist. *Stability and error bounds in the numerical integration of ordinary differential equations*, volume 130 of *Trans. of the Royal Inst. of Techn. Stockholm, Sweden.* Almqvist & Wiksells, Uppsala, Sweden, 1959. One citation on page 24.

G. Dahlquist. A special stability problem for linear multistep methods. *BIT Numer. Math.*, 3 (1):27–43, 1963. doi: 10.1007/BF01963532. 3 citations on pages 22, 51, and 54.

G. Dahlquist. On accuracy and unconditional stability of linear multistep methods for second order differential equations. *BIT Numer. Math.*, 18(2):133–136, 1978. doi: 10.1007/BF01931689. 2 citations on pages 51 and 54.

W.J.T. Daniel. Explicit/implicit partitioning and a new explicit form of the generalized alpha method. *Comm. Numer. Meth. Engrg.*, 19:909–920, 2003. doi: 10.1002/cnm.640. One citation on page 61.

Abaqus 6.11 – Theory manual. Dassault Systèmes (SIMULIA), 2011. downloaded from `http://abaqus.ethz.ch:2080/v6.11/pdf_books/THEORY.pdf` on September 23rd 2015. One citation on page 51.

P. Deuflhard. *Newton methods for nonlinear problems. Affine invariance and adaptve algorithms.* Springer Berlin Heidelberg New York, 2004. 2 citations on pages 110 and 115.

P. Deuflhard, E. Hairer, and J. Zugck. One-step and extrapolation methods for differential-algebraic systems. *Numer. Math.*, 51:501–516, 1987. doi: 10.1007/BF01400352. 3 citations on pages 73, 74, and 77.

W. E and B. Engquist. The heterogeneous multiscale methods. *Comm. Math. Sci.*, 1(1):87–132, 2003. doi: 10.4310/CMS.2003.v1.n1.a8. 2 citations on pages 26 and 125.

P. Eberhard and W. Schiehlen. Hierarchical modeling in multibody dynamics. *Arch. Appl. Mech.*, 68(3–4):237–246, 1998. doi: 10.1007/s004190050161. One citation on page 22.

D.G. Ebin. The motion of slightly compressible fluids viewed as a motion with strong constraining force. *Annals Math.*, 105(1):141–200, 1977. doi: 10.2307/1971029. One citation on page 37.

E. Eich-Soellner and C. Führer. *Numerical Methods in Multibody Dynamics.* B. G. Teubner Stuttgart, 1998. doi: 10.1007/978-3-663-09828-7. 6 citations on pages 7, 8, 15, 92, 110, and 113.

H. Elmqvist, S.E. Mattson, and M. Otter. Modelica – the new object-oriented modeling language. In R.N. Zobel. and D.P.F. Möller, editors, *Proc. of the 12th European Simulation Multiconference*, pages 127–131. SCS Europe, 1998. One citation on page 9.

S. Erlicher, L. Bonaventura, and O. Bursi. The analysis of the generalized-α method for non-linear dynamic problems. *Comput. Mech.*, 28:83–104, 2002. doi: 10.1007/s00466-001-0273-z. 14 citations on pages 49, 50, 53, 55, 57, 58, 65, 73, 76, 96, 106, 107, 112, and 113.

M. Etchechoury and C. Muravchik. Nonstandard singular perturbation systems and higher index differential-algebraic systems. *Appl. Math. Comp.*, 134(2–3):323–344, 2003. doi: 10.1016/S0096-3003(01)00288-0. 2 citations on pages 29 and 32.

J.E. Flaherty and R.E. O'Malley Jr. Analytical and numerical methods for nonlinear singular singularly perturbed initial value problems. *SIAM J. Appl. Math.*, 38:225–248, 1980. doi: 10.1137/0138020. One citation on page 29.

L. Fox and E.T. Goodwin. Some new methods for the numerical integration of ordinary differential equations. *Mathem. Proc. Cambridge Philosoph. Soc.*, 45(3):373–388, 1949. doi: 10.1017/S0305004100025007. One citation on page 48.

Chr. Fredebeul. A-BDF: A generalization of the backward differentiation formulae. *SIAM J. Numer. Anal.*, 35(5):1917–1938, 1998. doi: 10.1137/S0036142996306217. One citation on page 70.

C. Führer and B.J. Leimkuhler. Numerical solution of differential-algebraic equations for constrained mechanical motion. *Numer. Math.*, 59(1):55–69, 1991. doi: 10.1007/BF01385770. 2 citations on pages 16 and 18.

Function Bay Inc. RecurDyn - product brochure, downloaded on October 12th 2015. `http://functionbay.co.kr/pdf/brochure_e_web.pdf`. One citation on page 51.

Y.C. Fung and P. Tong. *Classical and Computational Solid Mechanics*, volume 1 of *Advanced Series in Engineering Science.* World Scientific Publishing Singapore New Jersey London Hong Kong, 2001. doi: 10.1142/4134. One citation on page 51.

A. Gallrein, M. Bäcker, M. Burger, and A. Gizatulin. An advanced flexible realtime tire model and its integration into Fraunhofer's driving simulator. Technical Report 2014-01-0861, SAE Technical Paper, 2014. One citation on page 60.

F.R. Gantmacher. *The theory of matrices.* Chelsea Publishing Company, New York, 1959. One citation on page 12.

C.W. Gear and D.R. Wells. Multirate linear multistep methods. *BIT Numerical Mathematics*, 24(4):484–502, 1984. doi: 10.1007/BF01934907. One citation on page 26.

C.W. Gear, B. Leimkuhler, and G. Gupta. Automatic integration of Euler–Lagrange equations with constraints. *J. Comput. Appl. Math.*, 12&13:77–90, 1985. doi: 10.1016/0377-0427(85) 90008-1. One citation on page 19.

M. Géradin and A. Cardona. *Flexible Multibody Dynamics – A finite element approach.* John Wiley & Sons, Ltd., 2001. 3 citations on pages 110, 111, and 112.

M. Géradin and D.J. Rixen. *Mechanical vibrations – Theory and application to structural dynamics.* John Wiley & Sons, Ltd., 3rd edition, 2015. 2 citations on pages 50 and 60.

I. Gladwell and R.M. Thomas. Stability properties of the Newmark, Houbolt and Wilson θ methods. *Int. J. Num. Analyt. Meth. Geomech.*, 4:143–158, 1980. doi: 10.1002/nag.1610040205. 3 citations on pages 52, 55, and 67.

A.A. Goldstein. Cauchy's method of minimization. *Numer. Math.*, 4:146–150, 1962. doi: 10.1007/BF01386306. One citation on page 115.

G.H. Golub and Ch.F. Van Loan. *Matrix Computations.* John Hopkins University Press, 3rd edition, 1996. One citation on page 67.

Z.-M. Gu, N.N. Nefedov, and R.E. O'Malley Jr. On singular singularly perturbed initial value problems. *SIAM J. Appl. Math.*, 49(1):1–25, 1989. doi: 10.1137/0149001. One citation on page 29.

E. Hairer. Unconditionally stable methods for second order differential equations. *Numer. Math.*, 32(4):373–379, 1979. doi: 10.1007/BF01401041. 2 citations on pages 54 and 55.

E. Hairer and G. Wanner. *Solving ordinary differential equations part II - Stiff and differential-algebraic problems.* Springer, 2nd edition, 2002. doi: 10.1007/978-3-642-05221-7. 16 citations on pages 7, 9, 12, 13, 17, 19, 21, 23, 24, 29, 50, 68, 70, 73, 78, and 117.

E. Hairer, Chr. Lubich, and M. Roche. Error of Runge–Kutta methods for stiff problems studied via differential algebraic equations. *BIT Numer. Math.*, 28:678–700, 1988. doi: 10.1007/BF01941143. One citation on page 44.

E. Hairer, Chr. Lubich, and M. Roche. *The numerical solution of differential-algebraic systems by Runge–Kutta-methods*, volume 1409 of *Lecture Notes in Mathematics.* Springer, 1989a. doi: 10.1007/BFb0093947. 4 citations on pages 12, 37, 44, and 114.

E. Hairer, Chr. Lubich, and M. Roche. Error of Rosenbrock methods for stiff problems studied via differential algebraic equations. *BIT Numer. Math.*, 29:77–90, 1989b. doi: 10.1007/BF01932707. One citation on page 70.

E. Hairer, S.P. Nørsett, and G. Wanner. *Solving ordinary differential equations part I - Nonstiff problems.* Springer Berlin Heidelberg New York, 2nd edition, 1993. doi: 10.1007/978-3-540-78862-1. 6 citations on pages 21, 54, 55, 56, 69, and 99.

Th. Hans. *Interaktive Simulation biomechanischer Bewegungsabläufe.* PhD thesis, Eberhard Karls University Tübingen, 2004. 3 citations on pages 7, 31, and 44.

I. Higueras. Numerical methods for stiff index 3 DAEs. *Math. Comput. Model. Dyn. Sys.*, 7(2): 239–262, 2001. doi: 10.1076/mcmd.7.2.239.3648. One citation on page 30.

H.M. Hilber and Th.J.R. Hughes. Collocation, dissipation and 'overshoot' for time integration schemes in structural dynamics. *Earthq. Eng. Struct. Dyn.*, 6:99–117, 1978. doi: 10.1002/eqe.4290060111. 3 citations on pages 50, 59, and 65.

H.M. Hilber, Th.J.R. Hughes, and R.L. Taylor. Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthq. Eng. Struct. D.*, 5:99–118, 1977. doi: 10.1002/eqe.4290050306. 3 citations on pages 47, 59, and 64.

C. Hoff and P.J. Pahl. Development of an implicit method with numerical dissipation from a generalized single-step algorithm for structural dynamics. *Comput. Meth. Appl. Mech. Engrg.*, 67:367–385, 1988a. doi: 10.1016/0045-7825(88)90053-9. 6 citations on pages 49, 55, 59, 64, 65, and 113.

C. Hoff and P.J. Pahl. Practical performance of the θ_1 method and comparison with other dissipative algorithms in structural dynamics. *Comput. Meth. Appl. Mech. Engrg.*, 67:87–110, 1988b. doi: 10.1016/0045-7825(88)90070-9. 4 citations on pages 22, 47, 65, and 113.

Th.J.R. Hughes. *The finite element method: Linear static and dynamic finite element analysis.* Prentice-Hall, Inc. Englewood Cliffs, NJ, 1987. 10 citations on pages 7, 11, 25, 26, 42, 47, 49, 56, 62, and 63.

Th.J.R. Hughes and W.K. Liu. Implicit-explicit finite elements in transient analysis: stability theory. *J. Appl. Mech.*, 45(2):371–374, 1978. doi: 10.1115/1.3424304. One citation on page 60.

G.M. Hulbert and J. Chung. The unimportance of spurious roots of time integration algorithms for structural dynamics. *Comm. Numer. Meth. Engrg.*, 10:591–597, 1994. doi: 10.1002/cnm.1640100803. One citation on page 55.

G.M. Hulbert and J. Chung. Explicit time integration algorithms for structural dynamics with optimal numerical dissipation. *Comput. Meth. Appl. Mech. Engrg.*, 137:175–188, 1996. doi: 10.1016/S0045-7825(96)01036-5. 3 citations on pages 50, 60, and 69.

M. Jahnke, K. Popp, and B. Dirr. Approximate analysis of flexible parts in multibody systems using the finite element method. In W. Schiehlen, editor, *Advanced Multibody System Dynamics*, volume 20 of *Solid Mechanics and Its Applications*, pages 237–256. Springer Netherlands, 1993. doi: 10.1007/978-94-017-0625-4_12. One citation on page 26.

K.E. Jansen, Chr.J. Whiting, and G.M. Hulbert. A generalized-α method for integrating the filtered Navier-Stokes equations with a stabilized finite element method. *Comp. Meth. Appl. Mech. Engrg.*, 190:305–319, 2000. doi: 10.1016/S0045-7825(00)00203-6. 4 citations on pages 48, 61, 70, and 112.

L.O. Jay. Structure preservation for constrained dynamics with super partitioned additive Runge–Kutta methods. *SIAM J. Sci. Comput.*, 20:416–446, 1999. doi: 10.1137/S1064827595293223. One citation on page 67.

L.O. Jay. Convergence of the generalized-α method for constrained systems in mechanics with nonholonomic constraints. Technical Report 181, University of Iowa, 2011. 2 citations on pages 50 and 94.

L.O. Jay and D. Negrut. Extensions of the HHT-α-method to differential-algebraic equations in mechanics. *Electron. T. Numer. Ana.*, pages 190–208, 2007. 4 citations on pages 50, 72, 77, and 118.

L.O. Jay and D. Negrut. A second order extension of the generalized-α method for constrained systems in mechanics. In C. Bottasso, editor, *Multibody Dynamics. Computational Methods and Applications*, volume 12 of *Computational Methods in Applied Sciences*, pages 143–158. Springer, 2008. doi: 10.1007/978-1-4020-8829-2_8. 3 citations on pages 64, 72, and 77.

C.T. Kelley. *Iterative methods for linear and nonlinear equations.* Number 16 in Frontiers in Applied Mathematics. SIAM Philadelphia, 1995. doi: 10.1137/1.9781611970944. 2 citations on pages 114 and 115.

M. Kettmann (Köbis). Das Generalized-α-Verfahren: Stabilität und Erweiterung auf Index-3-DAEs. Bachelor's thesis, Martin Luther University Halle-Wittenberg, 2009. 6 citations on pages 53, 59, 72, 76, 96, and 118.

M.A. Köbis and M. Arnold. Convergence of generalized-α time integration for nonlinear systems with stiff potential forces. In S.-S Kim and J.H. Choi, editors, *Proc. of the 3rd IMSD and the 7th ACMD, BEXCO, Busan, Korea*, pages 461–462 (ext. abstract), full paper: 10 pages, June/July 2014. 2 citations on pages 101 and 105.

M.A. Köbis and M. Arnold. Convergence of generalized-α time integration for nonlinear systems with stiff potential forces. *Multibody Syst. Dyn.*, 37(1):107–125, 2016. doi: 10.1007/s11044-015-9495-2. 2 citations on pages 101 and 105.

F. Kramer. Convergence results for linear multistep methods applied to quasi-singular perturbed problems, 2006. extended abstract 5-th International conference Aplimat, Bratislava (Slovakia), `https://www.kfs.oeaw.ac.at/publications/2006_Aplimat_Abstract_LMM_for_QSPP.pdf`. One citation on page 32.

P. Kunkel and V. Mehrmann. *Differential-algebraic equations: Analysis and numerical solution.* EMS Publishing House Zürich, 2006. doi: 10.4171/017. One citation on page 12.

A.J. Kurdila, J.L. Junkins, and S. Hsu. Lyapunov stable penalty methods for imposing holonomic constraints in multibody system dynamics. *Nonlin. Dyn.*, 4(1):51–82, 1993. doi: 10.1007/BF00047121. 4 citations on pages 37, 38, 44, and 45.

Th. Kurz, P. Eberhard, Chr. Henninger, and W. Schiehlen. From Neweul to Neweul-M^2: symbolical equations of motion for multibody system analysis and synthesis. *Multibody Syst. Dyn.*, 24:25–41, 2010. doi: 10.1007/s11044-010-9187-x. software version of December 21st 2010 (latest changes), see `http://www.itm.uni-stuttgart.de/research/neweul/neweulm2_de.php`. One citation on page 51.

Bryansk State Technical University (Russia) Laboratory of Computational Mechanics. Universal mechanism version 7.0 – user manual, downloaded October 12th 2015. `http://www.universalmechanism.com/download/70/eng/04_um_simulation_program.pdf`. One citation on page 51.

J. Lang and J.G. Verwer. ROS3P – an accurate third-order Rosenbrock solver designed for parabolic problems. *BIT Numer. Math.*, 41(4):731–738, 2001. doi: 10.1023/A:1021900219772. One citation on page 70.

B. Leimkuhler, L.R. Petzold, and C.W. Gear. Approximation methods for the consistent initialization of differential-algebraic equations. *SIAM J. Numer. Anal.*, 28(1):205–226, 1991. doi: 10.1137/0728011. 2 citations on pages 15 and 25.

A. Lew, J.E. Marsden, M. Ortiz, and M. West. Variational time integrators. *Int. J. Numer. Meth. Engrg.*, 60(1):153–212, 2004. doi: 10.1002/nme.958. One citation on page 26.

W.M. Lionen, J.J.B. de Swart, and W.A. van der Veen. Test set for IVP solvers. Technical Report NM-R9615, Centrum voor Wiskunde en Informatica (CWI), Department of Numerical Mathematics, 1996. 2 citations on pages 117 and 118.

Livermore Software Tech.-Corp. *LS-DYNA User's Manual (Keywords/Theory)*, 2006/2007. downloaded from `www.lstc.com/pdf/ls-dyna_971_manual_k.pdf`, October 25th 2015 and `http://www.lstc.com/pdf/ls-dyna_theory_manual_2006.pdf`, November 3rd 2015. One citation on page 51.

P. Lötstedt. On a penalty function method for the simulation of mechanical systems subject to constraints. Technical Report TRITA-NA-7919, Dept. of Num. Anal. & Comp. Sci. Stockholm, 1979. One citation on page 38.

Chr. Lubich. Extrapolation integrators for constrained multibody systems. *Impact Comput. Sci. Engrg.*, 3:213–234, 1991. doi: 10.1016/0899-8248(91)90008-I. 2 citations on pages 20 and 111.

Chr. Lubich. Integration of stiff mechanical systems by Runge–Kutta methods. *Z. Angew. Math. Phys.*, 44:1022–1053, 1993. doi: 10.1007/BF00942763. 12 citations on pages 27, 37, 38, 39, 40, 43, 44, 114, 115, 116, 122, and 124.

Chr. Lubich and D. Weiss. Numerical integrators for motion under a strong constraining force. *Multiscale Model. Simul.*, 12(4):1592–1606, 2014. doi: 10.1137/14096092X. One citation on page 26.

Chr. Lubich, U. Nowak, U. Pöhle, and Chr. Engstler. MEXX – Numerical software for the integration of constrained mechanical multibody systems. Technical Report SC 92–12, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 1992. 2 citations on pages 8 and 111.

D.G. Luenberger. Dynamic equations in descriptor form. *IEEE Trans. on Autom. Control*, AC-22(3):312–321, 1977. doi: 10.1109/TAC.1977.1101502. One citation on page 14.

Chr. Lunk and B. Simeon. Solving constrained mechanical systems by the family of Newmark and α-methods. *Z. Angew. Math. Mech.*, 86:772–784, 2006. doi: 10.1002/zamm.200610285. 5 citations on pages 50, 72, 73, 77, and 118.

A.A. Medovikov. High order explicit methods for parabolic equations. *BIT Numer. Math.*, 38 (2):372–390, 1998. doi: 10.1007/BF02512373. One citation on page 25.

N. Minorsky. *Nonlinear oscillations.* Litton Educational Publishing, Inc., 1962. reprint by R. E. Krieger Pub. Co., 1987. One citation on page 26.

K. Modin and G. Söderlind. Geometric integration of Hamiltonian systems perturbed by Rayleigh damping. *BIT Numer. Math.*, 51(4):977–1007, 2011. doi: 10.1007/s10543-011-0345-1. One citation on page 45.

MSC Software Corp. *MSC Adams Manual (About Adams/Solver) and MSC Nastran Dynamic Analysis User's Guide*, downloaded October 5th 2015. `https://simcompanion.mscsoftware.com`. One citation on page 51.

A. Müller and Z. Terze. On the choice of configuration space for the numerical Lie group integration of constrained rigid body systems. *J. Comput. Appl. Math.*, 262:3–13, 2014. doi: 10.1016/j.cam.2013.10.039. One citation on page 92.

K. Nachbagauer, K. Sherif, and W. Witteveen. FreeDyn – a multibody simulation research code. In E. Oñate, X. Oliver, and A. Huerta, editors, *Proc. of 11th World Congress on Comput. Mech. (WCCM), 5th Europ. Conf. Comput. Mech. (ECCM) and 6th Europ. Congress Comp. Fluid Dyn. (ECFD), Barcelona, Catalonia, Spain.* CIMNE Center for Numer. Meth. Engrg., 2015. One citation on page 51.

D. Negrut, R. Rampalli, G. Ottarson, and A. Sajdak. On the use of the HHT method in the context of index 3 differential algebraic equations of multibody dynamics. In *Proc. of ASME: 5th International Conference on Multibody Systems, Nonlinear Dynamics, and Control, Long Beach CA*, volume 6, pages 207–218, 2005. doi: 10.1115/DETC2005-85096. 4 citations on pages 51, 64, 72, and 113.

N.M. Newmark. A method of computation for structural dynamics. *J. Eng. Mech. Div.*, 85(EM 3):67–94, 1959. Proc. of ASCE. 4 citations on pages 47, 48, 58, and 65.

K. Nipp. Numerical integration of differential algebraic systems and invariant manifolds. *BIT Numerical Mathematics*, 42(2):408–439, 2002. doi: 10.1023/A:1021959227466. 2 citations on pages 34 and 92.

W. Nolting. *Grundkurs Theoretische Physik 2 – Analytische Mechanik.* Springer Berlin Heidelberg, 7th edition, 2006. doi: 10.1007/978-3-642-41980-5. One citation on page 10.

H. Olsson, H. Elmqvist, and M. Otter. *Modelica – A unified object-oriented language for systems modeling, language specification, version 3.3.* Modelica Association, 2012. downloaded from `https://www.modelica.org/documents/ModelicaSpec33.pdf` on May 23rd 2016, 20:00. One citation on page 9.

R.E. O'Malley Jr. *Singular Perturbation Methods for Ordinary Differential Equations.* Springer, 1991. doi: 10.1007/978-1-4612-0977-5. 2 citations on pages 27 and 29.

J.C.G. Orden and I. Romero. Energy-entropy-momentum integration of discrete thermo-visco-elastic dynamics. *Europ. J. Mech. A/Solids*, 32:76–87, 2012. doi: 10.1016/j.euromechsol.2011.09.007. One citation on page 71.

B. Owren and H.H. Simonsen. Alternative integration methods for problems in structural dynamics. *Comp. Meth. Appl. Mech. Engrg.*, 122(1–2):1–10, 1995. doi: 10.1016/0045-7825(94)00717-2. One citation on page 70.

C.C. Pantelides. The consistent initialization of differential-algebraic systems. *SIAM J. Sci. Stat. Comput.*, 9(4):213–231, 1988. doi: 10.1137/0909014. One citation on page 12.

L.R. Petzold. An efficient numerical method for highly oscillatory ordinary differential equations. *SIAM J. Numer. Anal.*, 18(3):455–479, 1981. doi: 10.1137/0718030. One citation on page 26.

L.R. Petzold. Differential/algebraic equations are not ODE's. *SIAM J. Sci. Stat. Comput.*, 3 (3):367–384, 1982. doi: 10.1137/0903023. 2 citations on pages 12 and 114.

L.R. Petzold and P. Lötstedt. Numerical solution of nonlinear differential equations with algebraic constraints II: practical implications. *SIAM J. Sci. Stat. Comput.*, 7(3):720–733, 1986. doi: 10.1137/0907049. One citation on page 114.

L.R. Petzold, L.O. Jay, and J. Yen. Numerial solution of highly oscillatory ordinary differential equations. *Acta Numerica*, 6:437–483, 1997. doi: 10.1017/S0962492900002750. 3 citations on pages 25, 29, and 54.

F.A. Potra and W.C. Rheinboldt. On the numerical solution of Euler–Lagrange equations. *Mech. Struct. & Mach.*, 19(1):1–18, 1991. doi: 10.1080/08905459108905135. One citation on page 18.

A. Prothero and A. Robinson. On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations. *Math. Comp.*, 28(125):145–162, 1974. doi: 10.1090/S0025-5718-1974-0331793-2. 3 citations on pages 68, 95, and 102.

J.W. Prüß, R. Schnaubelt, and R. Zacher. *Mathematische Modelle in der Biologie – Deterministische homogene Systeme.* Birkhäuser Basel Boston Berlin, 2008. doi: 10.1007/978-3-7643-8437-1. One citation on page 56.

J. Rang. Adaptive timestep control for the generalised-α method. In J.P. Moitinho de Almeida, P. Díez, C. Tiago, and N. Parés, editors, *VI Conference on Adaptive Modeling and Simulations, ADMOS 2013*, 2013. 3 citations on pages 48, 50, and 64.

S. Reich. Smoothed dynamics of highly oscillatory Hamiltonian systems. *Physica D*, 89:28–42, 1995. doi: 10.1016/0167-2789(95)00212-X. Postprint published at the institutional repository of Potsdam University. 3 citations on pages 28, 37, and 39.

W.C. Rheinboldt. Solution fields of nonlinear equations and continuation methods. *SIAM J. Numer. Anal.*, 17(2):221–237, 1980. doi: 10.1137/0717020. One citation on page 115.

W.C. Rheinboldt. Differential-algebraic systems as differential equations on manifolds. *Math. Comp.*, 43:473–482, 1984. doi: 10.1090/S0025-5718-1984-0758195-5. One citation on page 18.

W.C. Rheinboldt and B. Simeon. Computing smooth solutions of DAEs for elastic multibody systems. *Comput. Math. Appl.*, 37:69–83, 1999. doi: 10.1016/S0898-1221(99)00077-2. One citation on page 30.

H. Rubin and P. Ungar. Motion under a strong constraining force. *Comm. Pur. Appl. Math.*, 10:65–87, 1957. doi: 10.1002/cpa.3160100103. One citation on page 38.

Samtech. SAMCEF Mecano, Product Description, downloaded October 5th 2015. `http://www.suri.co.jp/news/pdf/4-SAMCEF_Mecano_en.pdf` and `http://www.plm.automation.siemens.com/en_us/products/lms/samtech/`. One citation on page 51.

G.G. Sanborn, J. Choi, and J.H. Choi. Review of RecurDyn integration methods. In S.-S Kim and J.H. Choi, editors, *Proc. of the 3rd IMSD and the 7th ACMD, BEXCO, Busan, Korea*, pages 359–360 (ext. abstract), full paper: 8 pages, June/July 2014. 2 citations on pages 51 and 65.

J.A. Sanders, F. Verhulst, and J. Murdock. *Averaging methods in nonlinear dynamical systems*. Number 59 in Applied Mathematical Sciences. Springer New York, 2nd edition, 2007. doi: 10.1007/978-0-387-48918-6. One citation on page 26.

J.M. Sanz-Serna and M.P. Calvo. *Numerical Hamiltonian problems*. Number 7 in Appl. Math. and Math. Comp. Chapman & Hall, 1994. One citation on page 10.

M. Schaub. *Numerische Integration steifer mechanischer Systeme mit impliziten Runge–Kutta-Verfahren*. PhD thesis, Technical University of Munich, 2004. One citation on page 68.

M. Schaub and B. Simeon. Automatic h-scaling for the efficient time integration of stiff mechanical systems. *Multibody Syst. Dyn.*, 8:329–345, 2002. doi: 10.1023/A:1020973630828. 2 citations on pages 95 and 114.

M. Schaub and B. Simeon. Blended Lobatto methods in multibody dynamics. *Z. Angew. Math. Mech.*, 83(10):720–728, 2003. doi: 10.1002/zamm.200310069. One citation on page 67.

W. Schiehlen. *Multibody system handbook*. Springer New York Heidelberg, 1990. doi: 10.1007/978-3-642-50995-7. 2 citations on pages 8 and 117.

St. Schneider. Convergence results for multistep Runge–Kutta methods on stiff mechanical systems. *Numer. Math.*, 69:495–508, 1995. doi: 10.1007/s002110050105. 3 citations on pages 44, 58, and 108.

S. Scholz. Order barriers for the B-convergence of ROW methods. *Computing*, 41(3):219–235, 1989. doi: 10.1007/BF02259094. 3 citations on pages 44, 70, and 102.

H.-R. Schwarz. Ein Verfahren zur Stabilitätsfrage bei Matrizen-Eigenwertproblemen. *Z. angew. Math. Phys.*, 7(6):473–500, 1956. doi: 10.1007/BF01601178. One citation on page 56.

L.F. Shampine. Implementation of Rosenbrock methods. *ACM Trans. Math. Softw.*, 8(2):93–113, 1982. doi: 10.1145/355993.355994. One citation on page 70.

L.F. Shampine and M.W. Reichelt. The Matlab ODE suite. *SIAM J. Sci. Comp.*, 18(1):1–22, 1997. doi: 10.1137/S1064827594276424. One citation on page 41.

E. Shchepakina, V. Sobolev, and M.P. Mortell. *Singular Perturbations – Introduction to system order reduction methods with applications*, volume 2114 of *Lecture Note in Mathematics*. Springer Cham Heidelberg New York Dordrecht London, 2014. doi: 10.1007/978-3-319-09570-7. One citation on page 29.

B. Siciliano and W.J. Book. A singular perturbation approach to control of lightweight flexible manipulators. *Int. J. Robot. Res.*, 7(4):79–90, 1988. doi: 10.1177/027836498800700404. One citation on page 25.

B. Simeon. Modelling a flexible slider crank mechanism by a mixed system of DAEs and PDEs. *Math. Model. Sys.*, 2(1):1–18, 1996. doi: 10.1080/13873959608837026. 2 citations on pages 117 and 118.

B. Simeon. Order reduction of stiff solvers at elastic multibody systems. *Appl. Numer. Math.*, 28:459–475, 1998. doi: 10.1016/S0168-9274(98)00060-9. 6 citations on pages 44, 65, 95, 102, 103, and 114.

B. Simeon. *Computational flexible multibody dynamics – a differential-algebraic approach.* Differential-algebraic equations forum. Springer Heidelberg New York Dordrecht London, 2013. doi: 10.1007/978-3-642-35158-7. 10 citations on pages 19, 27, 30, 39, 42, 68, 70, 91, 95, and 114.

B. Simeon. On the history of differential-algebraic equations – A retrospective with personal side trips, 2015. version of June 15th 2015, downloaded from `https://kluedo.ub.uni-kl.de/frontdoor/index/index/docId/4106` on September 24th 2015. One citation on page 1.

B. Simeon, R. Serban, and L.R. Petzold. A model of macroscale deformation and microvibration in skeletal muscle tissue. *ESAIM: Mathem. Model. Numer. Anal.*, 43:805–823, 2009. doi: 10.1051/m2an/2009030. 2 citations on pages 7 and 43.

J.C. Simo and N. Tarnow. The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics. *Z. Angew. Math. Phys.*, 43(5):757–793, 1992. doi: 10.1007/BF00913408. One citation on page 26.

G. Söderlind, L.O. Jay, and M. Calvo. Stiffness 1952–2012: Sixty years in search of a definition. *BIT Numer. Math.*, 55:531–558, 2015. doi: 10.1007/s10543-014-0503-3. One citation on page 24.

G. Steinebach and P. Rentrop. An adaptive method of lines approach for modelling flow and transient transport in rivers. In A. Vande Wouwer, Ph. Saucez, and W.E. Schiesser, editors, *Adaptive method of lines*, pages 181–205. Chapman & Hall/CRC, 2001. `http://www.mathworks.com/matlabcentral/fileexchange/10354-rodasp/content/rodasp.m`, December 10th 2015, 10:08. One citation on page 70.

K. Strehmel and R. Weiner. Linearly implicit Runge–Kutta methods and their modification for stiff problems. *Teubner-Texte zur Mathematik*, 107:288–294, 1989. One citation on page 70.

K. Strehmel, R. Weiner, and H. Podhaisky. *Numerik gewöhnlicher Differentialgleichungen.* Springer Spektrum, 2nd edition, 2012. One citation on page 56.

Th. Stumpp. *Integration stark gedämpfter mechanischer Systeme mit Runge–Kutta-Verfahren.* PhD thesis, Eberhard Karls University Tübingen, 2004. 3 citations on pages 33, 36, and 115.

Th. Stumpp. Integration of strongly damped mechanical systems by Runge–Kutta methods. In A. Bucchianico, R.M.M. Mattheij, M.A. Peletier, H.-G. Bock, F. Hoog, A. Friedman, W. Langford, H. Neunzert, W.R. Pulleyblank, T. Rusten, and A.-K. Tornberg, editors, *Progress in Industrial Mathematics at ECMI 2004*, volume 8 of *Mathematics in Industry*, pages 642–646. Springer Berlin Heidelberg, 2006. doi: 10.1007/3-540-28073-1_99. full material (21 pages) published as technical report, from `www.tat.physik.uni-tuebingen.de/sfb/reports/stumpp_207.ps.gz`, 2014. 2 citations on pages 35 and 124.

Th. Stumpp. Asymptotic expansions and attractive invariant manifolds of strongly damped mechanical systems. *Z. Angew. Math. Mech.*, 88:630–643, 2008. doi: 10.1002/zamm.200700057. 3 citations on pages 27, 32, and 33.

F. Takens. Motion under the influence of a strong constraining force. In *Global theory of dynamical systems: Proceedings of an international conference held at Northwestern University, Evanston, Illinois, June 18-22, 1979*, volume 819 of *Lecture Notes in Mathematics*, pages 425–445. Springer Berlin Heidelberg, 1980. doi: 10.1007/BFb0087006. One citation on page 38.

P.J. van der Houwen and B.P. Sommeijer. Explicit Runge–Kutta–Nyström methods with reduced phase errors for computing oscillating solutions. *SIAM J. Numer. Anal.*, 24:595–617, 1987. doi: 10.1137/0724041. One citation on page 102.

N.G. van Kampen and J.J. Lodder. Constraints. *Amer. J. Phys.*, 52:419–424, 1984. doi: 10.1119/1.13647. One citation on page 30.

A.B. Vasil'eva and V.F. Butuzov. Singularly perturbed equations in the critical case. Technical Report 2039, Mathematics Research Center, University of Wisconsin, Madison, 1980. One citation on page 29.

St. Vater, R. Klein, and O. M. Knio. A scale-selective multilevel method for long-wave linear acoustics. *Acta Geophysica*, 59(6):1076–1108, 2011. doi: 10.2478/s11600-011-0037-x. 2 citations on pages 70 and 121.

St. Weber, M. Arnold, and M. Valášek. Quasistatic approximations of stiff second order differential equations. *Appl. Numer. Math.*, 62(10):1579–1590, 2012. doi: 10.1016/j.apnum.2012.06.030. One citation on page 30.

R.A. Wehage and E.J. Haug. Generalized coordinate partitioning for dimension reduction in analysis of constrained dynamic systems. *J. Mech. Des*, 104(1):247–255, 1982. doi: 10.1115/1.3256318. One citation on page 17.

P. Wolfe. Convergence conditions for ascent methods. *SIAM Rev.*, 11(2):226–235, 1969. doi: 10.1137/1011036. One citation on page 115.

W.L. Wood, M. Bossak, and O.C. Zienkiewicz. An alpha modification of Newmark's method. *Int. J. Numer. Meth. Eng.*, 5:1562–1566, 1981. doi: 10.1002/nme.1620151011. 2 citations on pages 47 and 59.

X. Yan. Singularly perturbed differential/algebraic equations I: Asymptotic expansion of outer solutions. *J. Math. Anal. Appl.*, 207:326–344, 1997. doi: 10.1006/jmaa.1997.5256. One citation on page 30.

J. Yen and L.R. Petzold. An efficient Newton-type iteration for the numerical solution of highly oscillatory constrained multibody dynamic systems. *SIAM J. Sci. Comput.*, 19(5):1513–1534, 1998. doi: 10.1137/S1064827596297227. 2 citations on pages 30 and 115.

J. Yen, L.R. Petzold, and S. Raha. A time integration algorithm for flexible mechanism dynamics: The DAE α-method. *Comput. Meth. Appl. Mech. Engrg.*, 158(3–4):341–355, 1998. doi: 10.1016/S0045-7825(97)00261-2. 3 citations on pages 73, 77, and 115.

O.C. Zienkiewicz. A new look at the Newmark, Houbolt and other time stepping formulas. a weighted residual approach. *Earthq. Engrg. Struct. Dyn.*, 5(4):413–418, 1977. doi: 10.1002/eqe.4290050407. One citation on page 49.

List of Figures

Selbständigkeitserklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertation selbständig und ohne fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht.

Halle (Saale), den 30. Juni 2016

Lebenslauf

	Markus Arthur Köbis (geb. Kettmann) geboren am 30.09.1985 in Halle (Saale)
seit 02/16	Berater bei der ORSOFT GmbH Leipzig
10/10 – 01/16	wissenschaftlicher Mitarbeiter am Institut für Mathematik der Martin-Luther-Universität Halle–Wittenberg
01/12	Master of Science in Mathematik
11/09 – 01/12	Masterstudium der Mathematik Martin-Luther-Universität Halle–Wittenberg
11/09	Bachelor of Science in Mathematik
09/06 – 11/09	Bachelorstudium Mathematik mit Anwendungsfach (Physik) Martin-Luther-Universität Halle–Wittenberg
07/05 – 03/06	Grundwehrdienst
07/05	Abitur
07/96 – 07/05	Südstadt-Gymnasium Halle (Saale)