CISM International Centre for Mechanical Sciences 565
Courses and Lectures

Peter Betsch *Editor*

Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics

International Centre
for Mechanical Sciences

Springer

CISM International Centre for Mechanical Sciences

Courses and Lectures

Volume 565

The series presents lecture notes, monographs, edited works and proceedings in the field of Mechanics, Engineering, Computer Science and Applied Mathematics. Purpose of the series is to make known in the international scientific and technical community results obtained in some of the activities organized by CISM, the International Centre for Mechanical Sciences.

More information about this series at http://www.springer.com/series/76

Peter Betsch
Editor

Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics

Springer

Editor
Peter Betsch
Institute of Mechanics
Karlsruhe Institute of Technology
Karlsruhe
Germany

ISSN 0254-1971 ISSN 2309-3706 (electronic)
CISM International Centre for Mechanical Sciences
ISBN 978-3-319-31877-6 ISBN 978-3-319-31879-0 (eBook)
DOI 10.1007/978-3-319-31879-0

Library of Congress Control Number: 2016935608

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This volume contains notes based on lectures presented at the advanced course 'Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics' held at the International Centre for Mechanical Sciences (CISM) in Udine, Italy, during October 7–11, 2013.

The objective of the five chapters in this volume is to provide insight into state-of-the-art numerical methods for nonlinear structural and flexible multibody dynamics. In the field of structural mechanics, finite element methods are commonly applied for the discretization in space. Due to the large dimension of the resulting semi-discrete system, one is typically content with second-order accurate schemes for the discretization in time.

Based on well-established time-stepping schemes for the linear regime, energy-momentum consistent schemes and energy dissipating variants thereof have been developed in the framework of nonlinear structural dynamics during the past 25 years. These schemes are known to possess superior numerical stability and robustness properties when compared to standard methods.

The chapter written by I. Romero provides a general overview of high-frequency dissipative integrators for linear and nonlinear elastodynamics. If the controllable numerical dissipation is switched off, one typically gets back to energy-momentum consistent schemes that are addressed in the chapter authored by P. Betsch.

Due to the presence of finite rotations, the configuration space of multibody systems is typically nonlinear. In the chapter written by M. Arnold, A. Cardona, and O. Brüls, Lie group integrators are presented which preserve the Lie group structure of the underlying nonlinear configuration space by design.

An alternative route to the design of structure-preserving numerical methods are variational integrators. The newly emerging class of variational integrators is the topic of the chapter authored by A.J. Lew and P. Mata A. Last but not least the Chapter written by J. Gerstmayr, A. Humer, P. Gruber, and K. Nachbagauer provides insight into the absolute nodal coordinate formulation which is increasingly popular in the field of flexible multibody dynamics.

The combination of these chapters provides a unique perspective on up-to-date numerical methods for nonlinear structural dynamics and flexible multibody dynamics. Sincere thanks are due to the colleagues for preparing their chapters for this volume. Special thanks to Professors Martin Arnold, Alberto Cardona, Johannes Gerstmayr, Adrian Lew, and Ignacio Romero for taking part at the course and presenting their excellent lectures.

The course brought together nearly 40 participants from 8 countries. We are grateful to all participants for their interest and the numerous discussions that took place during and after the lectures. We are particularly thankful to the Scientific Council of CISM for supporting this course and recognizing the importance of the topic. We further thank the CISM staff for the excellent organization, support, and hospitality. Professor Paolo Serafini is gratefully acknowledged for his encouragement to publish these lecture notes and his patience to wait for the final versions.

Peter Betsch

Contents

High Frequency Dissipative Integration Schemes for Linear and Nonlinear Elastodynamics

Ignacio Romero

Abstract Time integration schemes with controllable, artificial, high frequency dissipation are extremely common in practical engineering analyses for integrating in time initial boundary value problems previously discretized in space with finite elements or similar techniques. In this chapter, we describe the structure of the most commonly employed integration schemes of this type and focus in their numerical analysis for linear and nonlinear problems. These include spectral, energy, and backward error analyses. For the nonlinear case, additionally, we study the preservation of conservation laws and the approximation of relative equilibria. The chapter should provide a general overview of dissipative methods, their issues, and the tools available for their formulation and analysis.

1 Introduction

Stiff ordinary differential equations, such as the ones commonly appearing in solid dynamics and many other areas of applied mathematics, have solutions that involve characteristic times of very different orders of magnitude. Whereas one is often interested only in the slow response, integrating the fastest time scales is sometimes necessary and always dictated by the time step choice (Wood 1990; Hairer and Wanner 1991).

In systems of ordinary differential equations resulting from the spatial discretization of partial differential equations, the modes with highest frequencies are inevitably resolved very poorly by the mesh. This is the case, for example, in solid mechanics, where a finite element mesh—or a similar discretization technique—is

I. Romero (✉)
Department of Mechanical Engineering, Technical University of Madrid,
Madrid, Spain
e-mail: ignacio.romero@upm.es

I. Romero
IMDEA Materials Institute, Getafe, Madrid, Spain

P. Betsch (ed.), *Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics*, CISM International Centre for Mechanical Sciences 565, DOI 10.1007/978-3-319-31879-0_1

employed to approximate the initial boundary value problem of continuum elastodynamics by a *semidiscrete* initial value problem, governed by the same equations as the problem of structural dynamics, which can then be integrated in time numerically. In such a process, the one of interest in this chapter, the highest frequencies modes are completely spurious, and so poorly resolved that their precise value is often irrelevant for the analyst.

The mathematical analysis and the numerical experience accumulated during decades indicates that, in nonlinear problems, the poorly resolved, high frequency modes of the solution are ultimately responsible for many instabilities observed in the numerical solution of stiff evolution problems. Since, as already mentioned, those same modes are poorly resolved when deriving from a spatial discretization, time integration algorithms that possess some kind of high frequency controllable dissipation are frequently favored in research and commercial codes. It is the goal of this chapter to discuss in which sense this is a valid approach and how it should be addressed from the standpoint of the user and the algorithmic designer.

To understand the strengths and limitations of high frequency dissipative integrators for solid mechanics, it is convenient to start by studying them in the context of linear elastodynamics. The equations that describe this problem are amenable to a complete mathematical analysis and guide the choice of algorithms that can later be applied to more complex nonlinear problems. Most of the efforts in this regard have been addressed toward the development of direct integration schemes, similar to Newmark's classical method (Newmark 1956), but with optimal dissipation properties and maximum accuracy. In fact, some members of the Newmark family of methods, possibly the most commonly used integrators in solid and structural dynamics, have controllable high frequency dissipation, although all of these are only first-order accurate (Hughes 1983). The design of Newmark-like integrators for solid and structural dynamics with second-order accuracy and controllable high frequency dissipation motivated a large amount of works since the 1960s (Wilson 1968; Bathe and Wilson 1973; Hilber et al. 1977; Wood et al. 1981; Bazzi and Anderheggen 1982; Zienkiewicz et al. 1984; Chung and Hulbert 1993; Modak and Sotelino 2002; Zhou and Tamma 2004). Many commercial finite element and multibody codes have adopted one of these methods as the default integrator for implicit dynamical problems.

The stability of a time integration method, when employed in the solution of linear elastodynamics, is best understood when a complete spectral analysis is performed (Hughes 1983, 1987; Wood 1990; Bathe 1996). Such an analysis characterizes the evolution in time of each of the independent modes that contribute to the global solution, identifying their growth or decay, phase error, overshoot, etc. Alternatively, a more direct method of analysis based on the energy of the solution can be employed to assess the properties of the integrators (Hughes 1976, 1983; Romero 2002, 2004). The latter approach, although less systematic than the spectral analysis, furnishes global information that completes the information obtained from the former.

The analysis of time integration schemes for general, nonlinear problems in elastodynamics demands completely different techniques. Since spectral analyses cannot be performed on nonlinear equations, the study of the stability of time integrators is often limited to the assessment of the energy evolution (Belytschko and Schoeberle 1975; Simo and Wong 1991; Simo and Tarnow 1992, 1994; Betsch and Steinmann 2001; Kuhl and Crisfield 1999; Armero and Romero 2001a, b; Bathe 2007). A noteworthy exception is the work of Erlicher et al. (2002), where a complete nonlinear analysis of the Generalized-α method is presented. In the context of nonlinear problems, however, stability analysis is not the only interesting information one needs to know about a time integration scheme. Given that the dynamics of conserving—and dissipative—solid mechanics often possesses symmetries and conservation laws it is desirable to employ time integration schemes which can preserves *exactly* as many invariants as possible, even when controllable high frequency dissipation is introduced in the integration. With this goal in mind, many works have tried to extend classical high frequency dissipative schemes. See, among others, the works by Bauchau and Theron (1996), Kuhl and Ramm (1996), Bottasso and Borri (1997), Kuhl and Ramm (1999), Bauchau and Joo (1999), Kuhl and Crisfield (1999), Armero and Petocz (1999), Armero and Romero (2001a, b), Bottasso and Borri (1997), Armero and Romero (2003).

In this chapter we study time stepping algorithms for linear and nonlinear solid elastodynamics, with a special emphasis in methods that possess controllable high frequency dissipation. These methods are widely employed in solid, structural, and multibody dynamics, but, for concreteness, we restrict our presentation to the former. It should be noted that multibody dynamical systems often include, in addition to rigid and flexible parts, a large number of constraints that turn the governing equations into differential algebraic systems. The analysis techniques required to study these are not the same as the ones presented in this chapter.

The chapter is structured as follows. First, the problem of linear elastodynamics in described in Sect. 2. Section 3 discusses the discretization of elastodynamics, including specific examples, and presents their numerical analysis based on spectral and energy methods, and the modified differential equation. In Sect. 4 the initial boundary problem of nonlinear elastodynamics is reviewed, with a special emphasis in conservation laws and relative equilibria. There are a large number of methods that have been specially designed for this problem, as described above, but in Sect. 5 we focus in the EDMC family of high frequency dissipative methods. We discuss again their numerical analysis and study the solution of relative equilibria, by the EDMC and other dissipative methods. Section 6 closes the chapter with a summary of results.

2 Linear Elastodynamics

We present in this section the problem of continuum linear elastodynamics and describe some of its main properties.

2.1 The Weak Form of the Equations of Linear Elastodynamics

The initial boundary value problem of linear elastodynamics is presented in weak form since finite elements, the only spatial discretization considered in this chapter, are based on the variational form of these equations. For that, consider a deformable body occupying a set $\mathcal{B} \subset \mathbb{R}^N$, where N can be 2 or 3, with boundary $\partial\mathcal{B}$ partitioned into disjoint subsets $\partial_u\mathcal{B}$ and $\partial_t\mathcal{B}$. Denoting points in $\mathcal{B}$ as $\boldsymbol{x}$, the displacement of one of them at time $t \in \mathcal{I} = [0, T]$ is given by $\boldsymbol{u}(\boldsymbol{x}, t)$. Using the notation $\dot{()}$ for the partial derivative with respect to time, the velocity and acceleration fields are, respectively, $\dot{\boldsymbol{u}}$ and $\ddot{\boldsymbol{u}}$.

Assuming that the solid is elastic and homogeneous with elasticity tensor $\mathbb{C}$, and density ρ, the initial boundary value problem of linear elastodynamics consists in finding $\boldsymbol{u}$ in the set

$$\begin{aligned}\mathcal{S} = \{\boldsymbol{u} : \mathcal{B} \times \mathcal{I} \to \mathbb{R}^N,\ \boldsymbol{u}(\cdot, t) \in [H^1(\mathcal{B})]^N,\\ \dot{\boldsymbol{u}}(\boldsymbol{x}, \cdot), \ddot{\boldsymbol{u}}(x, \cdot) \in [L_2(\mathcal{B})]^N\ ,\ \boldsymbol{u} = \boldsymbol{0} \text{ on } \partial_u\mathcal{B} \times \mathcal{I}\}\ ,\end{aligned} \tag{1}$$

such that, for every $\boldsymbol{v}$ in

$$\mathcal{V} = \{\boldsymbol{w} : \mathcal{B} \to \mathbb{R}^N, \boldsymbol{w} \in [H_1(\mathcal{B})]^N, \boldsymbol{w} = \boldsymbol{0} \text{ on } \partial_u\mathcal{B}\}, \tag{2}$$

the following variational equation is satisfied:

$$\int_{\mathcal{B}} \boldsymbol{\sigma} \cdot \varepsilon[\boldsymbol{w}]\, \mathrm{d}V + \int_{\mathcal{B}} \rho\, \dot{\boldsymbol{v}} \cdot \boldsymbol{w}\, \mathrm{d}V = \int_{\mathcal{B}} \boldsymbol{b} \cdot \boldsymbol{w}\, \mathrm{d}V + \int_{\partial_t\mathcal{B}} \boldsymbol{h} \cdot \boldsymbol{w}\, \mathrm{d}A\ \ , \tag{3}$$

with $\boldsymbol{v} = \dot{\boldsymbol{u}}$ and H^1 being the Hilbert space of square integrable functions with square integrable derivatives. In the previous equation, and below, $\boldsymbol{b}$ is a known field of body forces and $\boldsymbol{h}$ a known field of tractions on the set $\partial_t\mathcal{B}$; $\boldsymbol{\sigma} = \mathbb{C}\varepsilon$ is the stress tensor, and $\varepsilon = (\nabla\boldsymbol{u} + \nabla^T\boldsymbol{u})/2$ is the tensor of infinitesimal strain. In addition to (3), the solution $\boldsymbol{u}$ must satisfy the following initial conditions

$$\boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u}_o(\boldsymbol{x})\ ,\quad \dot{\boldsymbol{u}}(\boldsymbol{x}, 0) = \boldsymbol{v}_o(\boldsymbol{x})\ ,\quad \boldsymbol{x} \in \mathcal{B}\ , \tag{4}$$

where $\boldsymbol{u}_o$, $\boldsymbol{v}_o$ are the known initial displacement and velocity, respectively. The dot product in Eq. (3), and hereafter, denotes the inner product between tensors of any order resulting from the pairwise contraction of all their indices.

2.2 Energy

A pair $(\boldsymbol{u}, \boldsymbol{v})$ will be referred to as a state of the system, and will be denoted by the symbol z. Given any state z, its energy $E(z)$ is defined as the sum of kinetic and potential energy of a body with displacement $\boldsymbol{u}$ and velocity $\boldsymbol{v}$, i.e.,

$$E(z) = \frac{1}{2} \int_{\mathcal{B}} \varepsilon[\boldsymbol{u}] \cdot \mathbb{C}\varepsilon[\boldsymbol{u}] \, \mathrm{d}V \; + \frac{1}{2} \int_{\mathcal{B}} \rho \, \boldsymbol{v} \cdot \boldsymbol{v} \, \mathrm{d}V \;\; , \tag{5}$$

and its (energy) norm as

$$||z||_E = \sqrt{E(z)} \; . \tag{6}$$

The properties of the potential and kinetic energy guarantee that the function $|| \cdot ||_E$ has indeed all the properties of a norm. Moreover, it is a natural norm for the problem (3) which is reasonable for studying the convergence and stability of numerical approximations of elastodynamics.

The evolution of the energy along a solution to the problem (3) is obtained by taking the derivative of (5), using (3), and setting $\dot{\boldsymbol{u}} = \boldsymbol{v}$:

$$\begin{aligned} \dot{E} &= \int_{\mathcal{B}} \varepsilon[\boldsymbol{u}] \cdot \mathbb{C}\varepsilon[\dot{\boldsymbol{u}}] \, \mathrm{d}V \; + \int_{\mathcal{B}} \rho \, \dot{\boldsymbol{v}} \cdot \boldsymbol{v} \, \mathrm{d}V \\ &= \int_{\mathcal{B}} \boldsymbol{b} \cdot \boldsymbol{v} \, \mathrm{d}V \; + \int_{\partial_t \mathcal{B}} \boldsymbol{h} \cdot \boldsymbol{v} \, \mathrm{d}A \;\; , \end{aligned} \tag{7}$$

which corresponds to the external power exerted on the body. Obviously, if the external forces vanish, the energy of the system is conserved. This a priori estimate on the growth of the energy norm of the solution is the basis of the energy method of stability analysis.

3 Direct Integration Schemes for Linear Elastodynamics

We describe next the most commonly employed strategy for approximating numerically the solution to linear elastodynamics. The idea, often referred in the literature as the *method of lines* consists in projecting the solution in space first and later integrating in time the resulting ordinary differential equations (see, e.g., Belytschko 1983; Hughes 1987; Wood 1990; Zienkiewicz and Taylor 2005).

3.1 The Spatial Discretization

We proceed next to study the spatial discretization of (3) by means of finite elements. This is done for concreteness, since it is the most common approach in solid dynamics. However, almost all the results that follow are valid when the finite element spatial discretization is replaced by any other approximation method in space.

To define the finite element projection, let us consider a mesh on $\mathcal{B}$ connecting a set $\mathcal{N}$ of nodes. The mesh is assumed to be regular and h its mesh size parameter. The basic step in any Galerkin method is to replace the trial space $\mathcal{S}$ and the weighting space $\mathcal{V}$ by finite dimensional subsets of the form

$$\begin{aligned} \mathcal{S}^h &= \{\boldsymbol{u}^h(\boldsymbol{x},t) = \sum_{a\in\mathcal{N}} N^a(\boldsymbol{x})\,\mathsf{u}^a(t)\ ,\ \text{with } \mathsf{u}^a(t) \in [\mathcal{C}^2(\mathcal{I})]^N, \\ &\qquad \boldsymbol{u}^h(\boldsymbol{x},t) = \mathbf{0} \text{ on } \partial_u\mathcal{B}, \}\ , \\ \mathcal{V}^h &= \{\boldsymbol{w}^h(\boldsymbol{x}) = \sum_{a\in\mathcal{N}} N^a(\boldsymbol{x})\,\mathsf{w}^a,\quad \text{with } \mathsf{w}^a \in \mathbb{R}^N, \\ &\qquad \boldsymbol{w}^h(\boldsymbol{x}) = \mathbf{0} \text{ on } \partial_u\mathcal{B}\}\ , \end{aligned} \tag{8}$$

where $\mathcal{C}^2(\mathcal{I})$ is the space of continuous functions in $\mathcal{I}$, with continuous derivatives up to order 2. The shape functions $N^a : \mathcal{B} \to \mathbb{R}$ are piecewise polynomials with compact support and sufficient smoothness. The vectors $\mathsf{u}^a(t)$ are the nodal displacement functions and likewise, w^a are the nodal values of the weighting functions. The semidiscrete solution $\boldsymbol{u}^h$ is the displacement function in $\mathcal{S}^h$ that verifies (3) for every $\boldsymbol{w}^h$ in the the space $\mathcal{V}^h$. Denoting by n_{dof} the number of unknown nodal components of the displacement $\boldsymbol{u}^h$, the weak equation (3) can be written, after a standard manipulation, in matrix form as system of n_{dof} ordinary differential equations:

$$\begin{aligned} \mathsf{M}\ddot{\mathsf{U}}(t) + \mathsf{K}\mathsf{U}(t) &= \mathsf{f}(t), \qquad t \in \mathcal{I} \\ \mathsf{U}(0) &= \mathsf{U}_o \\ \dot{\mathsf{U}}(0) &= \mathsf{V}_o\ . \end{aligned} \tag{9}$$

In these equations, M is the mass matrix, K the stiffness matrix, f the vector of external forces and U the vector of nodal displacements, which collects in a single vector all the unknown nodal components u^a. The solution to the system of ordinary differential equations (9) is denoted $\boldsymbol{z}^h = (\boldsymbol{u}^h, \boldsymbol{v}^h)$ with $\boldsymbol{v}^h(\boldsymbol{x},t) = \dot{\boldsymbol{u}}(\boldsymbol{x},t)$. Abusing the notation, the same symbol will sometimes refer to the vector form of the semidiscrete displacement and velocity, i.e., $\boldsymbol{z}^h = (\mathsf{U}, \dot{\mathsf{U}})$.

We note that there is no damping in Eq. (9) because the constitutive law employed in its derivation is purely elastic. A damping term of the form $\mathsf{C}\dot{\mathsf{U}}$, where C is the damping matrix, might show in the dynamic equilibrium either due to a viscous contribution in the material response or added *ad hoc* due to, i.e., Rayleigh damping.

3.2 The Time Discretization

To complete the numerical approximation of the equations of motion, the system of ordinary differential equations (9) needs to be solved and a numerical method is used to integrate it. In this work we consider a certain type of linear multistep algorithms which includes the most commonly employed integrators for linear elastodynamics, namely Newmark's method, the HHT method, the Wilson's θ-method, Hulbert's α-method, and others. Before defining this class of methods, consider a partition of the time interval $\mathcal{I}$ into N subintervals $[t_n, t_{n+1}]$ of constant size $\Delta t = t_{n+1} - t_n$, where $0 = t_o < t_1 < \cdots < t_{N-1} < t_N = T$ and denote by $\boldsymbol{y}_n$ the numerical approximation at time t_n of a variable $\boldsymbol{y}$. With this notation, we have the following

Definition 3.1 (Geradin 1974; Hughes 1987) A linear $k-$step method for the second-order differential equation (9) is a rule of the form

$$\sum_{i=0}^{k} \left[\alpha_i \mathsf{M}\mathsf{U}_{n+i} + \Delta t^2 \beta_i \left(-\mathsf{K}\mathsf{U}_{n+i} + \mathsf{f}(t_{n+i})\right)\right] = \mathbf{0}\,, \tag{10}$$

together with the initial conditions

$$\mathsf{U}_i = \mathsf{U}(i\,\Delta t), \quad i = 1-k, \ldots, 0\,. \tag{11}$$

The scalars α_i, β_i are constants and define each method.

The initial conditions (11) are given at negative time instants for simplicity of notation and are equivalent to the usual initial conditions at positive time instants, up to a time shift.

Many integration schemes have been developed for structural and continuum elastodynamics, some of which have been mentioned in Sect. 1. As working examples we mention two of them:

Example 3.2 Newmark's method (Newmark 1956) is possibly the most commonly employed direct integration scheme for continuum and structural dynamics. Given the displacement U_n, velocity V_n and acceleration A_n at time t_n, the value of these vectors at time t_{n+1} is obtained by solving

$$\begin{aligned} &\mathsf{M}\mathsf{A}_{n+1} + \mathsf{C}\mathsf{V}_{n+1} + \mathsf{K}\mathsf{U}_{n+1} = \mathsf{f}_{n+1}\,, \\ &\mathsf{U}_{n+1} = \mathsf{U}_n + \Delta t \mathsf{V}_n + \frac{\Delta t}{2}\left((1-2\beta)\mathsf{A}_n + 2\beta\mathsf{A}_{n+1}\right)\,, \\ &\mathsf{V}_{n+1} = \mathsf{V}_n + \Delta t\left((1-\gamma)\mathsf{A}_n + \gamma\mathsf{A}_{n+1}\right)\,. \end{aligned} \tag{12}$$

In these equations (β, γ) are algorithmic parameters that select individual members of the family, each of them with different properties. We note, for example, that second order can only be attained if $\gamma = \frac{1}{2}$. A full analysis of this method can be found, e.g., in Hughes (1987).

For reasons that will become apparent later, certain perturbations of Newmark's method have favorable properties from the numerical standpoint. As an example of this class of integrators we consider the following:

Example 3.3 The HHT method (Hilber et al. 1977) is a one-parameter family of implicit algorithms for continuum and structural dynamics. As before, given $(\mathsf{U}_n, \mathsf{V}_n, \mathsf{A}_n)$, the approximations to the displacement, velocity and acceleration, respectively, at time t_n, their corresponding values at time t_{n+1} are obtained from the system of equations:

$$\begin{aligned} &\mathsf{M}\mathsf{A}_{n+1} + \mathsf{C}\mathsf{V}_{n+\alpha} + \mathsf{K}\mathsf{U}_{n+\alpha} = \mathsf{f}_{n+\alpha} \ , \\ &\mathsf{U}_{n+1} = \mathsf{U}_n + \Delta t \mathsf{V}_n + \frac{\Delta t}{2}\left((1-2\beta)\mathsf{A}_n + 2\beta \mathsf{A}_{n+1}\right) \ , \\ &\mathsf{V}_{n+1} = \mathsf{V}_n + \Delta t\left((1-\gamma)\mathsf{A}_n + \gamma \mathsf{A}_{n+1}\right) \ , \\ &0.7 \le \alpha \le 1 \ , \quad \beta = \left(1 - \frac{\alpha}{2}\right)^2 \ , \quad \gamma = \frac{3}{2} - \alpha \ . \end{aligned} \tag{13}$$

With the notation employed, α is the only free parameter and, as shown in Hilber et al. (1977), it selects among different members of the family, all being second-order accurate.

3.3 Spectral Properties of Direct Integration Schemes

As noted above, the most commonly used integration schemes in elastodynamics are linear multistep methods for the second-order equation (9) and therefore may be formulated as in Eq. (10). For analysis purposes, it proves convenient to express these methods as a single-step recurrence relation. For example, if the velocity and acceleration at time t_n are denoted, respectively, as V_n, A_n, linear 3-step methods for elastodynamics can be expressed as

$$\mathcal{Z}_{n+1} = \mathcal{A}\mathcal{Z}_n + \mathcal{B}_n \ , \qquad \text{with} \qquad \mathcal{Z}_n = \left(\mathsf{U}_n, \Delta t\, \mathsf{V}_n, \Delta t^2\, \mathsf{A}_n\right)^T \ . \tag{14}$$

The so-called amplification matrix $\mathcal{A}$ has dimensions $(3n_{dof})^2$, and depends on the parameters of the method; the vector $\mathcal{B}_n$ has length $3n_{dof}$ and is obtained from a weighted evaluation of the forcing terms at times t_{n-1}, t_n and t_{n+1}.

Since (14) is a *linear* recurrence equation, much information can be inferred from the spectral properties of the amplification matrix alone. But, moreover, since the semidiscrete problem itself is defined by *linear* differential equations, the time evolution of each of the modes of the solution is independent from the rest, and it suffices to study the properties of the *modal* amplification matrices. More details on this simplification can be found, for instance, in Hughes (1987). Following the example above, the modal recurrence relation for a linear 3-step method must be of the form

$$\mathcal{Z}^\omega_{n+1} = \mathcal{A}_\omega \mathcal{Z}^\omega_n + \mathcal{B}^\omega_n \qquad \text{with} \qquad \mathcal{Z}^\omega_n = \left(u^\omega_n, \Delta t v^\omega_n, \Delta t^2 a^\omega_n\right)^T \ . \tag{15}$$

In this equation, the vector $\mathcal{Z}_n^{\omega}$ holds the amplitude, rate, and acceleration at time t_n and t_{n+1} of the mode with frequency ω. The matrix $\mathcal{A}_{\omega}$ and vector $\mathcal{B}_n^{\omega}$ are, respectively, the amplification matrix of the $\omega-$mode and the forcing associated with this mode corresponding to the interval $\mathcal{I}_n$.

Expression (15) is very convenient for the study of the algorithm. It characterizes in a very simple and compact way the evolution of each mode in the solution. In particular, this frequency analysis helps to understand the dissipation (or lack thereof) in the integration of the modes. Even though direct integration schemes do not make use of any modal decomposition, they treat each mode in the solution independently, as relation (15) indicates.

Since the equations of elastodynamics are linear, an explicit expression for the modal amplification matrix can always be found, which would be a function of the method's parameters and the nondimensional frequency $\Omega = \omega \Delta t$. Once this expression is obtained, all the conservation/dissipation properties of the method can be deduced solely from its spectrum $\{\lambda_i\}$. In particular, the *spectral radius* ρ of the matrix or its *algorithmic damping ratio* ξ, defined respectively as

$$\rho = \max(|\lambda_i|), \qquad \xi = -\frac{1}{\Omega} \log(\max(|\lambda_i|)) \;, \tag{16}$$

measure the growth or decay of each individual mode in a single step. Spectral stability requires the spectral value not to be larger than one, and that eigenvalues of unit modulus be simple. In addition, the spectral radii of dissipative integration schemes need to be smaller than 1 for high frequencies.

Example 3.4 Following our previous example, the modal amplification matrix of the HHT method can be shown to be

$$\mathcal{A} = \begin{bmatrix} \Omega^2\alpha & 0 & 1 \\ 1 & 0 & -\beta \\ 0 & 1 & -\gamma \end{bmatrix}^{-1} \begin{bmatrix} \Omega^2(\alpha-1) & 0 & 0 \\ 1 & 1 & \frac{1}{2}-\beta \\ 0 & 1 & 1-\gamma \end{bmatrix} . \tag{17}$$

The spectral radius of this matrix and its algorithmic damping ratio, for three values of the algorithmic parameter α, is depicted in Fig. 1. It is clearly observed that the spectral radius is never greater than 1 and that for $\alpha < 1$, it decreases monotonically with Ω. It is ρ_∞, the limit value for $\Omega \to \infty$, the one that defines the algorithmic treatment of the high frequencies. An algorithm with efficient high frequency dissipation must possess $\rho_\infty < 1$, and controllable by the use via a parameter choice. In the case of the HHT method, the parameter α clearly modulates the amount of high frequency dissipation. The same conclusion can be drawn from the plot of the algorithmic damping ratio.

If the case of Newmark's method, it is possible to have high frequency algorithmic dissipation, as Fig. 2 shows, but only at the expense of selecting $\gamma > \frac{1}{2}$, i.e., by losing the second-order accuracy. Methods such as HHT were developed to preserve the second-order accuracy of the trapezoidal rule (Newmark's method with $(\beta, \gamma) = (\frac{1}{4}, \frac{1}{2})$, even with the addition of dissipation.

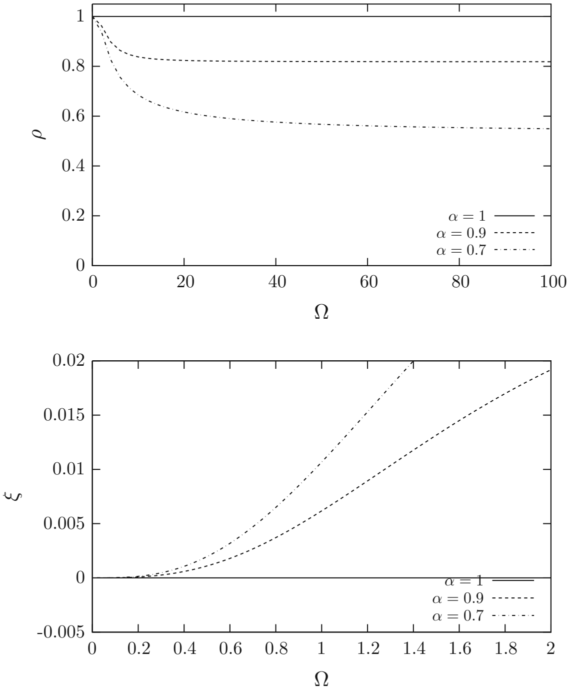


Fig. 1 Spectral radius (*top*) and algorithmic damping ratio (*bottom*) of the HHT method for three values of the parameter α (1, 0.9, and 0.7)

3.4 Energy Stability Analysis

A space-time discretization is unconditionally stable when, for a forcing free problem, the total energy in the solution is *uniformly* bounded by the initial energy, i.e.,

$$\|(\mathsf{U}(t), \dot{\mathsf{U}}(t)\|_E \leq C\|(\mathsf{U}_o, \mathsf{V}_o)\|_E \ , \tag{18}$$

for some constant C independent of h and Δt. This is the discrete counterpart of the a priori energy estimate that was derived for the continuum problem and forms the basis for the *energy method* of stability analysis.

It is often argued that spectral analyses, as described in Sect. 3.3, can give necessary and sufficient conditions for stability of time stepping methods in linear elasticity. The spectral stability condition, namely that the spectral radius must be smaller than one for all frequencies or strictly smaller than one in the case of repeated eigenvalues, is equivalent to the energy-boundedness of each of the modes. However, following Romero (2002, 2004), we claim that this is not enough to guarantee *uniform* boundedness across all modes in a solution and a stronger notion of stability must be established. As a corollary, it turns out that some spectrally stable methods such as the HHT are not unconditionally stable in the sense introduced before.

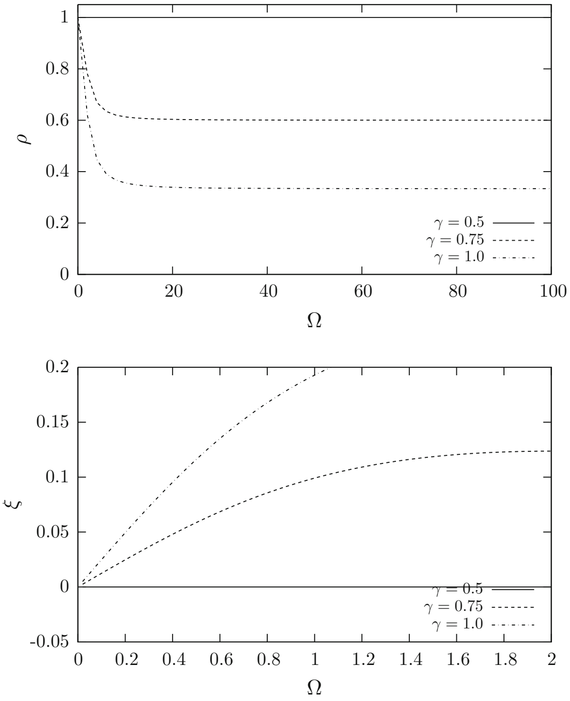


Fig. 2 Spectral radius (*top*) and algorithmic damping ratio (*bottom*) of Newmark's method for three values of the parameter γ (0.5, 0.75, 1.0) and the parameter $\beta = (\gamma/2 + 1/4)^2$ selected for maximum dissipation of high frequencies

The main result of Romero (2002, 2004) is the following:

Theorem 3.5 *A time stepping method employed in the time discretization of an initial boundary value problem is unconditionally stable when the amplification matrices $\mathcal{A}^{\omega_i}$ are spectrally stable and the set*

$$\mathcal{F} = \left\{ \mathcal{A}^{\omega}, \omega \in \mathbb{R}^+ \right\}$$

is compact.

For *one* system of ordinary differential equations, like the one in Eq. (9), the classical spectral stability criterion is necessary and sufficient for energy stability, since the set of all amplification matrices is finite, and thus compact. However, when a method is analyzed for all possible spatial discretizations of a problem, and in the limit when the number of degrees of freedom goes to infinity, one must study if there is a *uniform* bound for the energy, and Theorem 3.5 provides a sufficient condition.

3.5 Backward Stability Analysis

Spectral analysis has traditionally been the main technique to evaluate the dissipation properties of time integration schemes for elastodynamics, as described in Sect. 3.3. Next, we outline an alternative approach, seldom employed in the literature, that may cast some additional light on the dissipative/conservative behavior of these methods.

The idea of backward error analysis is to find a perturbed system whose exact evolution closely resembles the solution furnished by a given numerical method. Then, by analyzing the partial or ordinary differential equations of the perturbed system, one might obtain relevant information about the numerical method itself. This approach was first employed in the context of partial differential equations (see, e.g., Warming and Hyett 1974). Since then, this technique has been extensively employed for the stability analysis of linear and nonlinear problems (cf., Hairer 1994, 1999; Leimkuhler and Reich 2004, and references therein). A major advantage of this approach is that the conclusions obtained are valid for both the linear and nonlinear ranges.

We will use this idea to study time integrators, i.e., solutions to the semidiscrete problem of elastodynamics. If we consider a numerical scheme which is order m accurate, the goal of this analysis is to obtain a *perturbed differential equation* of linear elastodynamics, for which the given numerical scheme is order n accurate (with $n > m$). If we require that the solution of the modified differential equation exactly coincides at discrete points with the numerical solution, the former may be very difficult to obtain in general. However, enough information can be obtained if the time step size Δt is small and only the lower order terms of the modified equation are retained.

The procedure to obtain the modified equation is as follows. If a numerical scheme has a truncation error of the form $\tau = C_m \Delta t^m$ when it approximates a certain differential equation, there must exist another differential equation constructed by appending an $\mathcal{O}(\Delta t^m)$ term to the original one such that the resulting truncation error is one order higher, $\tilde{\tau} = C_{m+1} \Delta t^{m+1}$. To increase the accuracy, the process is repeated.

To clarify these concepts, a backward error analysis of the HHT method previously described is presented next.

Example 3.6 For the one degree of freedom, elastodynamic equation with no physical damping and no external forces ($\ddot{d} + \omega^2 d = 0$), the fourth order accurate modified differential equation corresponding to the HHT method is

$$\begin{aligned} &\ddot{d} + G_2 \omega^4 \Delta t^3 \dot{d} + (1 + G_1 \omega^2 \Delta t^2) \omega^2 d = 0, \qquad \text{with} \\ &G_1 = \frac{3}{4}\alpha^2 - \alpha + \frac{1}{12}, \qquad \text{and} \quad G_2 = \frac{1}{4}(1-\alpha)\alpha^2 \,. \end{aligned} \tag{19}$$

Figure 3 depicts G_1 and G_2 as functions of the parameter α.

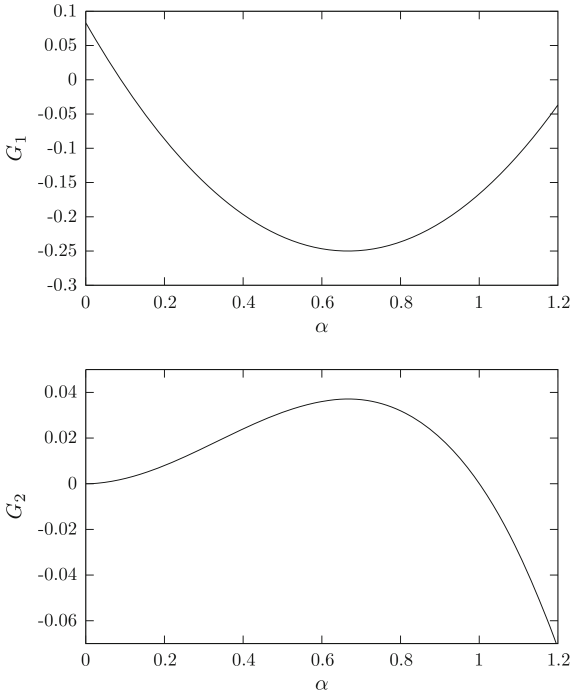


Fig. 3 Values of G_1 and G_2 in the modified equation of the HHT method

To prove this claim, let $\mathcal{A}$ be the amplification matrix (17) of the HHT method and let $(\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3)$ be its three principal invariants. Using Cayley-Hamilton's theorem we can write:

$$\mathcal{A}^3 - \mathcal{I}_1\mathcal{A}^2 + \mathcal{I}_2\mathcal{A} - \mathcal{I}_3\boldsymbol{I} = 0\ . \tag{20}$$

Let $z_n = (d_n, \Delta t v_n, \Delta t^2 a_n)^T$ be the vector of solution variables at time t_n. Multiplying both sides of Eq. (20) by the vector z_{n-2}, and using $z_{n+1} = \mathcal{A}^3 z_{n-2}$, $z_n = \mathcal{A}^2 z_{n-2}$, $z_{n-1} = \mathcal{A} z_{n-2}$ results in

$$z_{n+1} - \mathcal{I}_1 z_n + \mathcal{I}_2 z_{n-1} - \mathcal{I}_3 z_{n-2} = 0\ . \tag{21}$$

Finally, reordering the first row of this equation and calculating the expressions for the matrix invariants, the following 3-step formula is obtained

$$\frac{d_{n+1} - 2d_n + d_{n-1}}{\Delta t^2} + \frac{\omega^2}{D} d_n - \mathcal{I}_3 \frac{d_n - 2d_{n-1} + d_{n-2}}{\Delta t^2} = 0\ , \tag{22}$$

where $D = 1 + (1 - \frac{\alpha}{2})^2 \alpha \Omega^2$, $\mathcal{I}_3 = \frac{1}{4D}\alpha^2(\alpha - 1)\Omega^2$. The next step consists in inserting the Taylor expansions of d_{n+1}, d_n, d_{n-1} and d_{n-2} at time $t_{n+\alpha}$ in expression (22). After simplifying the resulting equations we obtain:

$$
\begin{aligned}
0 &= \ddot{d}(t_{n+\alpha}) + \omega^2 d(t_{n+\alpha}) \\
&+ \Delta t^2 \left[(\frac{1}{12} + \frac{\alpha^2}{2}) d^{(4)}(t_{n+\alpha}) + \alpha(1 - \frac{\alpha}{4})\omega^2 \ddot{d}(t_{n+\alpha}) \right] \\
&+ \Delta t^3 \left[(\frac{5}{6}\alpha - \frac{5}{4})\alpha^2 \omega^2 d^{(3)}(t_{n+\alpha}) - (\frac{1}{2} + \alpha^2)\frac{\alpha}{6} d^{(5)}(t_{n+\alpha}) \right] \\
&+ \mathcal{O}(\Delta t^4) \ .
\end{aligned}
\tag{23}
$$

Inserting $\ddot{d} = -G_2\omega^4\Delta t^3\dot{d} - (1 + G_1\omega^2\Delta t^2)\omega^2 d$ and its derivatives in Eq. (23) we observe that the $\mathcal{O}(\Delta t^2)$ and $\mathcal{O}(\Delta t^3)$ terms cancel for the choice

$$
G_1 = \frac{3}{4}\alpha^2 - \alpha + \frac{1}{12} \quad \text{and} \quad G_2 = \frac{1}{4}(1 - \alpha)\alpha^2 \ , \tag{24}
$$

which are precisely of the form indicated in Eq. (19).

The interest in the previous result is that some of the properties of the HHT method can be deduced by analyzing the solution to the *differential* equation (19).

First, we can interpret the modified differential equation as the equation of motion of a one degree of freedom oscillator of mass $m = 1$, spring constant $k = (1 + G_1\omega^2\Delta t^2)\omega^2$ and damping $c = G_2\omega^4\Delta t^2$. The natural frequency of the corresponding undamped system is $\sqrt{k/m} = \sqrt{1 + G_1\omega^2\Delta t^2}\omega$. In view of the value of the constant G_1 in (19), for $\alpha \in [0, 0.7]$, the algorithmic natural frequencies must be smaller than the exact ones. Hence, we expect that the periods in the numerically computed solution be longer than the exact ones.

Second, since the artificial damping introduced by the HHT is of the form $G_2\omega^4\Delta t^3$, for $\alpha \in [0, 0.7]$ this quantity is always nonnegative and maximum for $\alpha = 0.7$, as predicted by the spectral analysis (see Hilber et al. 1977). In particular, for $\alpha = 1$, no extra damping is added, as expected, since in this case the method reduces to the trapezoidal rule, which is conservative for linear problems. In addition, since the damping is proportional to Ω^4 we expect the numerical method to be dissipative for high frequency modes, with highest dissipation for the highest frequency modes.

Third, since for $\alpha > 1$ the constant G_2 is negative, the energy of the corresponding physical system will grow in time, even for vanishing forcing. This is clearly an unstable system and indicates that for $\alpha > 1$ the HHT method should be unstable, as observed in the original article of Hilber et al. (1977).

The previous analysis has been restricted to a one-dimensional, linear problem, but need not be. Although a full backward analysis of the method can provided insights not obtainable by the spectral analysis, this limited application serves to illustrate the connections between the two approaches.

Example 3.7 The following example shows graphically that the modified differential equation possesses an exact solution which closely resembles the numerical solution

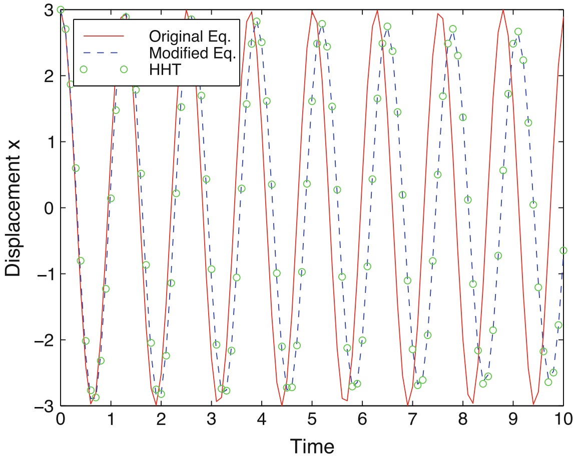


Fig. 4 Solutions to the exact and modified differential equations compared with the numerical solution obtained by the HHT method

obtained with the HHT method. In the following example the exact solutions to the differential equations

$$\ddot{d} + \omega^2 d = 0, \qquad d(0) = d_0, \quad v(0) = v_0 \ , \tag{25}$$

$$\ddot{d} + G_2\omega^4\Delta t^3\dot{d} + (1 + G_1\omega^2\Delta t^2)\omega^2 d = 0, \qquad d(0) = d_0, \quad v(0) = v_0 \ , \tag{26}$$

are compared with the numerical solution of Eq. (25) obtained with the HHT method. For this particular example, let $\omega = 5$, $\Delta t = 0.1$, $\alpha = 0.7$, $d_0 = 3$ and $v_0 = 0$. Figure 4 shows the exact solution to the differential equation (25), the solution to the modified differential equation (26), and the data points corresponding to the HHT solution of the former equation. It is clear that the solution to the modified equation and the HHT method are very close, illustrating that the modified equation effectively captures the correct dissipation and period error of the numerical scheme.

4 Nonlinear Elastodynamics

Following the same structure as for linear problems, we introduce in this section the problem of nonlinear continuum dynamics, focusing on the equations of the problem and the symmetries that it possesses. In the Sect. 5, we will present a family of high frequency dissipative methods for its solution, and we will discuss some of the difficulties entailed by the design of dissipative integrators for nonlinear problems.

Following standard notation in continuum mechanics (see, e.g., Gurtin 1981), let $\mathcal{B}_o \subset \mathbb{R}^N$ denote the reference configuration of a deformable body with points denoted by $\boldsymbol{X}$, and $\varphi(\cdot, t) : \mathcal{B}_o \to \mathbb{R}^N$ a one parameter family of deformations with $t \in [0, T]$ being the time. The boundary of the body admits the partition $\partial\mathcal{B}_o =$

$\partial_\varphi\mathcal{B}_o \cup \partial_t\mathcal{B}_o$. The (possibly empty) set $\partial_\varphi\mathcal{B}_o$ is such that $\boldsymbol{\varphi}(\cdot, t) = \bar{\boldsymbol{\varphi}}(t)$ on its points, with $\bar{\boldsymbol{\varphi}}$ a known function. Similarly, the tractions on the boundary subset $\partial_t\mathcal{B}_o \subset \partial\mathcal{B}_o$ are also known, and of value $\boldsymbol{T}(t)$. If the body is hyperelastic with stored energy function W and is subjected to body forces $\boldsymbol{B}$, its motion is described by an initial boundary value problem that, in weak form, is:

$$\int_{\mathcal{B}_o} \big(\boldsymbol{S} \cdot \boldsymbol{F}^T D[\delta\boldsymbol{\varphi}] \,\mathrm{d}V \,+ \rho_o\dot{\boldsymbol{V}} \cdot \delta\boldsymbol{\varphi}\,\mathrm{d}V \,- \rho_o\boldsymbol{B} \cdot \delta\boldsymbol{\varphi}\big)\,\mathrm{d}V \,= \int_{\partial_t\mathcal{B}_o} \boldsymbol{T} \cdot \delta\boldsymbol{\varphi}\,\mathrm{d}A \ , \tag{27}$$

for all admissible displacement variations $\delta\boldsymbol{\varphi}$, and $\boldsymbol{V} = \dot{\boldsymbol{\varphi}}$ being the material velocity. In these equations, and below, ρ_o is the reference density, $\boldsymbol{F} = D\boldsymbol{\varphi}$ is the deformation gradient, and $\boldsymbol{S}$, the second Piola–Kirchhoff stress tensor, is defined as $\boldsymbol{S} = 2\partial_{\boldsymbol{C}}W$ with $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}$ being the right Cauchy–Green deformation tensor.

4.1 Conservation Laws

The elastodynamic problem has a rich geometric structure that reveals itself more clearly when the problem is described in the Lagrangian or Hamiltonian formalism (Marsden and Hughes 1983; Simo et al. 1988). Keeping with the style of the presentation for linear systems we do not pursue either of these formalisms, discussed at length in some of the references provided, but recognize that the presence of symmetries in the Eq. (27) lead to conservation laws for the motion which are of great importance from the theoretical and practical points of view.

Before stating the most important conservation laws we define the linear momentum $\boldsymbol{L}$, the angular momentum $\boldsymbol{J}$ and the total energy H, respectively, as

$$\begin{aligned} \boldsymbol{L} &= \int_{\mathcal{B}_o} \rho_o\boldsymbol{V}\,\mathrm{d}V \ , \\ \boldsymbol{J} &= \int_{\mathcal{B}_o} \rho_o\boldsymbol{\varphi} \times \boldsymbol{V}\,\mathrm{d}V \ , \\ H &= \int_{\mathcal{B}_o} \Big(\frac{\rho}{2}|\boldsymbol{v}|^2 + W(\boldsymbol{\varphi})\Big)\,\mathrm{d}V \ . \end{aligned} \tag{28}$$

We summarize all the conservation laws in the following.

Theorem 4.1 *The motion of a body with $\partial_\varphi\mathcal{B}_o = \emptyset$ and no external forcing preserves the linear and angular momenta as well as the total energy.*

Proof The proof of these three conservation laws follows directly from Eq. (27) by choosing the admissible deformation variations to be, respectively, $\delta\boldsymbol{\varphi} = \boldsymbol{c}$, $\delta\boldsymbol{\varphi} = \boldsymbol{\varphi} \times \boldsymbol{c}$ and $\delta\boldsymbol{\varphi} = \boldsymbol{V}$ with $\boldsymbol{c}$ being a arbitrary constant vector field on $\mathcal{B}_o$. Details are omitted. □

4.2 Relative Equilibria

Another characteristic feature of nonlinear elastodynamics is the existence of a particular kind of solutions known as *relative equilibria* (Simo et al. 1991; Marsden and Ratiu 1994). These are motions along which not only the momenta and the energy are preserved, but also the deformation, as described for example by $\boldsymbol{C}$, is pointwise time-invariant. In these motions the body might translate and rotate, but without changing its shape. As shown in Simo et al. (1991) the boundary value problem that defines the relative equilibrium deformation $\boldsymbol{\varphi}_e$ in a reference frame attached to the center of mass of the solid for a given angular momentum $\boldsymbol{\mu}_e$ and given linear momentum $\boldsymbol{p}_e$ is

$$\begin{aligned} \boldsymbol{p}_e &= \rho_o \boldsymbol{\Omega} \times \varphi_e \\ \mathrm{DIV}[\boldsymbol{F}_e \boldsymbol{S}_e] &= \boldsymbol{\Omega} \times \boldsymbol{p}_e \end{aligned} \tag{29}$$

where $\boldsymbol{\Omega} = \boldsymbol{i}^{-1}(\varphi_e)\boldsymbol{\mu}_e$, and $\boldsymbol{i}(\varphi_e)$ is the inertia tensor at the equilibrium configuration.

The conservation laws stated in Theorem 4.1 and relative equilibria are important qualitative feature of the motion $\boldsymbol{\varphi}$. One of the driving stimulus for new discretization techniques is precisely the preservation of these features, without losing accuracy, or stability in the way.

5 Numerical Methods for Nonlinear Elastodynamics

As in the linear case, numerical methods for the equations of nonlinear continuum dynamics can be obtained in several ways. The most common of them is the method of lines already introduced in Sect. 3.1. As in the linear case, the deformation $\boldsymbol{\varphi}$ is approximated by a mapping $\boldsymbol{\varphi}^h \in \mathcal{S}^h$, with

$$\mathcal{S}^h = \left\{ \boldsymbol{\varphi}^h = \sum_{a \in \mathcal{N}} N^a(\boldsymbol{X}) \boldsymbol{\varphi}^a(t) \; , \boldsymbol{\varphi}^h = \bar{\boldsymbol{\varphi}} \text{ on } \partial_\varphi \mathcal{B}_o \right\} \; , \tag{30}$$

and the variations $\delta\boldsymbol{\varphi} \in \mathcal{V}^h$, with

$$\mathcal{V}^h = \left\{ \boldsymbol{w}^h = \sum_{a \in \mathcal{N}} N^a(\boldsymbol{X}) \boldsymbol{w}^a \; , \boldsymbol{w}^h = \boldsymbol{0} \text{ on } \partial_\varphi \mathcal{B}_o \right\} \; . \tag{31}$$

We note that the material velocities $\boldsymbol{V}^h$ belong to the same space as the deformation variations so their finite dimensional approximations will belong to $\mathcal{V}^h$ as well.

The time-continuous, spatially discrete version of the dynamic equilibrium equation is obtained by replacing the deformation, velocity and variations in Eq. (27)

by their counterparts in $\mathcal{S}^h$ and $\mathcal{V}^h$. To complete the discretization, again as in the linear case, the time-dependent fields are to be replaced by approximated values at the time instants $t_o, t_1, \dots, t_n$ and the rates by approximated values of the velocity and acceleration.

This strategy is very general, and one is free to select any of the time stepping methods mentioned in Sect. 1 to carry it out. We choose, however, to carry out the time discretization with a fairly general class of methods introduced in Armero and Romero (2001a, b), which are of the form

$$\begin{aligned}
\int_{\mathcal{B}_o} \left(\boldsymbol{S}^* \cdot \boldsymbol{F}^T_{n+1/2}\ D[\delta\boldsymbol{\varphi}^h] + \rho_o \frac{\boldsymbol{V}^h_{n+1} - \boldsymbol{V}^h_n}{\Delta t} \cdot \delta\varphi^h \right) \mathrm{d}V \\
= \int_{\mathcal{B}_o} \rho_o \boldsymbol{B}_{n+1/2} \cdot \delta\boldsymbol{\varphi}^h\ \mathrm{d}V + \int_{\partial_t \mathcal{B}_o} \boldsymbol{T}_{n+1/2} \cdot \delta\boldsymbol{\varphi}^h\ \mathrm{d}A \\
\int_{\mathcal{B}_o} \rho_o \boldsymbol{V}^* \cdot \delta\boldsymbol{V}^h\ \mathrm{d}V = \int_{\mathcal{B}_o} \rho_o \frac{\boldsymbol{\varphi}^h_{n+1} - \boldsymbol{\varphi}^h_n}{\Delta t} \cdot \delta\boldsymbol{V}^h\ \mathrm{d}V\ ,
\end{aligned} \tag{32}$$

with $\boldsymbol{S}^*$ and $\boldsymbol{V}^*$ being consistent approximations to the second Piola–Kirchhoff stress tensor and material velocity, respectively, at time $t_{n+1/2}$. More specifically, these two quantities are defined to be

$$\boldsymbol{S}^* = \boldsymbol{S}_{cons} + \boldsymbol{S}_{diss}\ , \qquad \boldsymbol{V}^* = \boldsymbol{V}_{cons} + \boldsymbol{V}_{diss}\ . \tag{33}$$

In the previous equation, the quantities denoted with $()_{cons}$ are the *conserving* part of the stress and velocity, respectively. They are exactly the same expressions of the midpoint stress and velocity of the Energy-Momentum method, as described in Gonzalez (2000), namely

$$\begin{aligned}
\boldsymbol{S}_{cons} &= 2(\mathbb{I} - \boldsymbol{N} \otimes \boldsymbol{N}) \frac{\partial W}{\partial \boldsymbol{C}}(\boldsymbol{C}_{n+1/2}) + \frac{W(\boldsymbol{C}_{n+1}) - W(\boldsymbol{C}_n)}{\|\boldsymbol{C}_{n+1} - \boldsymbol{C}_n\|}\boldsymbol{N} \\
&\quad \text{with } \boldsymbol{N} = \frac{\boldsymbol{C}_{n+1} - \boldsymbol{C}_n}{\|\boldsymbol{C}_{n+1} - \boldsymbol{C}_n\|}\ , \\
\boldsymbol{V}_{cons} &= \boldsymbol{V}^h_{n+1/2}\ .
\end{aligned} \tag{34}$$

To complete the method definition it remains to define the dissipative terms in Eq. (33), which must be of the form

$$\boldsymbol{S}_{diss} = f_{diss}\boldsymbol{N}\ , \qquad \boldsymbol{V}_{diss} = g_{diss}\boldsymbol{V}^h_{n+1/2}\ , \tag{35}$$

with f_{diss} and g_{diss} two scalar functions defined themselves as

$$f_{diss} = \frac{2\,\mathcal{D}_V}{\|\boldsymbol{C}_{n+1} - \boldsymbol{C}_n\|}\ , \qquad g_{diss} = \frac{2\,\mathcal{D}_K}{\|\boldsymbol{V}^h_{n+1}\|^2 - \|\boldsymbol{V}^h_n\|^2}\ . \tag{36}$$

The final step is the formulation of the terms $\mathcal{D}_V$ and $\mathcal{D}_K$, for which there is a certain freedom, as long as the consistency of the method is guaranteed and

$$\mathcal{D}_K + \mathcal{D}_V \geq 0\,. \tag{37}$$

This choice will be driven by the following result, which is the discrete counterpart of Theorem 4.1:

Theorem 5.1 *The scheme defined by Eqs.* (32)–(36) *preserves the conservation laws of momenta. Moreover, when there are no external forces applied on the body, its total energy is nonincreasing.*

Proof To show that the scheme preserves the conservation laws of the continuum, consider a nonlinear dynamical problem with no forcing and $\partial_\varphi \mathcal{B}_0 = \emptyset$. By choosing $\delta\boldsymbol{\varphi}^h = \boldsymbol{c}$, a constant field, in Eq. (32), it follows directly that

$$0 = \boldsymbol{c} \cdot \int_{\mathcal{B}_o} \rho_o \frac{\boldsymbol{V}^h_{n+1} - \boldsymbol{V}^h_n}{\Delta t}\, \mathrm{d}V \;=\; \frac{\boldsymbol{c}}{\Delta t} \cdot (\boldsymbol{L}_{n+1} - \boldsymbol{L}_n)\,, \tag{38}$$

which is just the discrete statement of the conservation of linear momentum. Similarly, by choosing $\delta\boldsymbol{\varphi}^h = \boldsymbol{c} \times \boldsymbol{\varphi}^h_{n+1/2}$, with $\boldsymbol{c}$ constant as before, we get

$$0 = \boldsymbol{c} \cdot \int_{\mathcal{B}_o} \rho_o \boldsymbol{\varphi}^h_{n+1/2} \times \frac{\boldsymbol{V}^h_{n+1} - \boldsymbol{V}^h_n}{\Delta t}\, \mathrm{d}V \;=\; \frac{\boldsymbol{c}}{\Delta t} \cdot (\boldsymbol{J}_{n+1} - \boldsymbol{J}_n)\,, \tag{39}$$

that proves the conservation of angular momentum. Finally, by selecting $\delta\boldsymbol{\varphi}^h = \boldsymbol{V}^h_{n+1/2}$ we obtain

$$\begin{aligned} 0 =& \frac{1}{\Delta t} \int_{\mathcal{B}_o} \boldsymbol{S}^* \cdot (\boldsymbol{C}_{n+1} - \boldsymbol{C}_n)\, \mathrm{d}V \;+ \int_{\mathcal{B}_o} \rho_o \boldsymbol{V}^* \cdot \frac{\boldsymbol{V}^h_{n+1} - \boldsymbol{V}^h_n}{\Delta t}\, \mathrm{d}V \\ =& \frac{1}{\Delta t} \left(H(\boldsymbol{\varphi}^h_{n+1}, \boldsymbol{V}^h_{n+1}) - H(\boldsymbol{\varphi}^h_n, \boldsymbol{V}^h_n) + \mathcal{D}_K + \mathcal{D}_V \right)\,. \end{aligned} \tag{40}$$

Hence,

$$H(\boldsymbol{\varphi}^h_n, \boldsymbol{V}^h_n) - H(\boldsymbol{\varphi}^h_{n+1}, \boldsymbol{V}^h_{n+1}) = \mathcal{D}_K + \mathcal{D}_V\,, \tag{41}$$

which is nonnegative, in view of Eq. (37). □

The previous theorem motivates that methods of the type described are referred to EDMC integrators, which stands for *Energy Dissipative, Momentum Conserving*. Depending on the expressions for the dissipative terms $\mathcal{D}_V$ and $\mathcal{D}_K$, first and second order methods have been proposed. In the first-order case, the EDMC-1 method, the expressions read:

$$\begin{aligned}\mathcal{D}_V &= \frac{\chi}{2}\left(\frac{1}{2}W(\boldsymbol{C}_{n+1}) + \frac{1}{2}W(\boldsymbol{C}_n) - W(\boldsymbol{C}_{n+1/2})\right),\\ \mathcal{D}_K &= \frac{\chi}{2}\left(\frac{1}{2}k(\boldsymbol{V}^h_{n+1}) + \frac{1}{2}k(\boldsymbol{V}^h_n) - W(\boldsymbol{V}^h_{n+1/2})\right),\end{aligned} \tag{42}$$

with $\chi \geq 0$ and $k(\boldsymbol{V}^h) = \frac{1}{2}\rho_o|\boldsymbol{V}^h|^2$ being the kinetic energy density. The parameter χ controls the size of the dissipation which grows proportionally to it. The EDMC-2 method, a second-order version of this type of integrators, is based on the expressions:

$$\begin{aligned}\mathcal{D}_V &= \left(\tilde{\boldsymbol{C}} - \boldsymbol{C}_n\right) \cdot \frac{1}{4}\mathbb{C}\,(\boldsymbol{C}_{n+1} - \boldsymbol{C}_n)\ ,\\ \mathcal{D}_K &= (\tilde{v} - v_n) \cdot \rho_o\,(v_{n+1} - v_n)\ ,\end{aligned} \tag{43}$$

with $v_{n+\alpha} = \|\boldsymbol{V}^h_{n+\alpha}\|$ and $\tilde{\boldsymbol{C}}$, $\tilde{v}$ two auxiliary variables implicitly defined by

$$\begin{aligned}\tilde{\boldsymbol{C}} &= (1-\beta)\boldsymbol{C}_n + \beta\boldsymbol{C}_{n+1}\\ \beta &= \alpha\frac{\Delta t}{h}(v_{n+1} - \tilde{v}_n)\\ \tilde{v}_n - v_n &= -\alpha\frac{\Delta t}{h}c^2(1-\tilde{\beta}_n)\|\boldsymbol{C}_{n+1} - \boldsymbol{C}_n\|^2\ ,\end{aligned} \tag{44}$$

for some user parameter α and a characteristic length and wave velocity denoted, respectively, as h and c. Details on the dissipative character and accuracy of the EDMC-2 method can be found in Armero and Romero (2001b).

Numerical integrators with artificial high frequency dissipation, like the ones discussed in Sect. 2 can be used, and are actually very frequently employed, for the solution of nonlinear problems. However, their greatest drawback is that they do not possess properties such as the ones proved in Theorem 5.1 for the EDMC integrators. First, most of these classical methods break the symmetries of the continuum problem, and often the conservation of angular momentum is spoiled. Second, the spectral dissipation, which is guaranteed in the linear setting cannot be proved in the nonlinear case, when the modes in the solution change as the solid deforms. On the contrary, it has been shown numerically that these methods can exhibit a pathological energy growth leading invariably to a solution blow-up (cf., e.g., Bauchau et al. (1995), Armero and Romero (2001a)). The conclusion is that the so-called unconditionally stable methods possess this property only for *linear* problems.

5.1 Relative Equilibria

As studied in Sect. 4.2, the equations of nonlinear continuum dynamics possess solutions in which, in addition to the usual conservation laws, the solid moves without changing its shape. From the numerical point of view it would be interesting to discern which integrators preserve these solutions.

It can be shown that any EDMC method, when given initial conditions in relative equilibrium, gives a solution which lies on the relative equilibrium for all times (see Armero and Romero 2001a). Moreover, we have the following stability result:

Theorem 5.2 *Given initial conditions sufficiently close to a relative equilibrium the solution obtained with an EDMC method will converge asymptotically to the latter.*

Proof It is shown in Simo et al. (1991) that the energy, as a function of deformation and velocity, has a minimum in the relative equilibria, when constrained to the level set of constant angular and linear momentum. Denoting by H_e this value, close to the relative equilibria, the function $\mathcal{L} = H - H_e$ is a Liapunov function. When an EDMC method is employed to solve a problem with initial conditions close to a relative equilibria, $\mathcal{L}$ will decrease monotonically, while on the level set of constant momenta, to its minimum, which corresponds to the relative equilibrium by construction. □

On the other hand, we show next that, for example, Newmark's method with $\gamma > \frac{1}{2}$ or the HHT method with $\alpha < 1$ can not preserve relative equilibria even for the simplest nonlinear problem, that of a point mass m attached to a fixed point through an elastic spring moving on a plane.

Let $\boldsymbol{z} = (\boldsymbol{q}, \boldsymbol{p}) \in \mathbb{R}^2 \times \mathbb{R}^2$ collect the position and momentum of the point mass on the plane and define the Hamiltonian of the system

$$H(\boldsymbol{z}) = V(\boldsymbol{q}) + K(\boldsymbol{p}) \,, \tag{45}$$

to be sum of the potential energy $V(\boldsymbol{q}) = \hat{V}(\lambda) = \frac{k}{2}(\lambda - \lambda_o)^2$, with $\lambda = |\boldsymbol{q}|$, and the kinetic energy $K = \frac{1}{2m}|\boldsymbol{p}|^2$. If the angular momentum of the system is $\boldsymbol{\mu}$, the equations that describe the relative equilibrium are:

$$\hat{V}'(\lambda_e) = \frac{|\boldsymbol{\mu}|^2}{m\lambda_e^3} \,, \qquad \pi_e = \frac{|\boldsymbol{\mu}|}{\lambda_e} \,, \tag{46}$$

with $\lambda_e = |\boldsymbol{q}_e|$, $\pi_e = |\boldsymbol{p}_e|$ and $\boldsymbol{\mu} = \boldsymbol{q} \times \boldsymbol{p}$. Moreover, this relative equilibrium is unique for every value of $|\mu|$ due to the convexity of $\hat{V}$.

From the property of conservation of angular momentum, both the position and momentum of the point mass must remain orthogonal to $\boldsymbol{\mu}$, and the relative equilibrium must be a motion in which the mass orbits around the fix point with constant spring length, equal to λ_e, and constant velocity $v_e = \pi_e/m$. To describe this motion mathematically we can consider rotations $\boldsymbol{Q} : \mathbb{R}^2 \to \mathbb{R}^2$ of the form

$$\boldsymbol{Q}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} . \tag{47}$$

The relative equilibria of the spring-mass system must be a motion

$$\begin{aligned} \boldsymbol{q}(t) &= \boldsymbol{Q}(\theta(t))\lambda_e \boldsymbol{e} , \\ \boldsymbol{p}(t) &= m\,\dot{\boldsymbol{Q}}(\theta(t))\lambda_e \boldsymbol{e} = \pi_e\, \boldsymbol{Q}(\theta(t))\, \boldsymbol{Q}(\pi/2)\boldsymbol{e} , \end{aligned} \tag{48}$$

for some unit vector $\boldsymbol{e} \in \mathbb{R}^2$.

We want to study next if the equations resulting from discretizing the equations of motion of the spring-mass system with an integrator admit solutions which are *discrete relative equilibria*, i.e., one parameter orbits of the point mass, and if they do, whether they fall on top of the exact trajectories, or not. More specifically, we would like to determine if an integrator can have solutions such as:

$$\begin{aligned} \boldsymbol{q}_n &= \boldsymbol{Q}_n \boldsymbol{q}_e = \boldsymbol{Q}_n \lambda_e \boldsymbol{e} , \\ \boldsymbol{p}_n &= \boldsymbol{Q}_n \boldsymbol{p}_e = \pi_e\, \boldsymbol{Q}_n\, \boldsymbol{Q}(\pi/2)\boldsymbol{e} , \end{aligned} \tag{49}$$

for some values of λ_e, π_e and $\boldsymbol{Q}_n = \boldsymbol{Q}(\theta_n)$. Define

$$\boldsymbol{Q}_{n+\alpha} = (1-\alpha)\,\boldsymbol{Q}_n + \alpha\,\boldsymbol{Q}_{n+1} , \qquad \boldsymbol{P} = \boldsymbol{Q}_{n+1}\,\boldsymbol{Q}_n^{-1} = \boldsymbol{Q}(\Delta\theta) , \tag{50}$$

where $\Delta\theta$ is the angle formed by $\boldsymbol{q}_n$ and $\boldsymbol{q}_{n+1}$. Note that $\boldsymbol{Q}_{n+\alpha}$ is not, in general, a rotation matrix.

The application of the HHT method, as given by Eq. (13), to the spring-mass model results, after eliminating the acceleration, in

$$\begin{aligned} (\boldsymbol{P}-\boldsymbol{I})\boldsymbol{q}_n &= \frac{\Delta t}{m}\boldsymbol{p}_n - m\nu\Big[(\frac{1}{2}-\beta)\boldsymbol{I} + \beta\boldsymbol{P}\Big]\boldsymbol{Q}_{n+\alpha}\boldsymbol{q}_n , \\ (\boldsymbol{P}-\boldsymbol{I})\boldsymbol{p}_n &= -\nu\frac{m}{\Delta t}\Big[(1-\gamma)\boldsymbol{I} + \gamma\boldsymbol{P}\Big]\boldsymbol{Q}_{n+\alpha}\boldsymbol{q}_n . \end{aligned} \tag{51}$$

with

$$\nu = \frac{\Delta t^2}{m}\frac{\hat{V}'(|\boldsymbol{q}_{n+\alpha}|)}{|\boldsymbol{q}_{n+\alpha}|} = \frac{\Delta t^2}{m}\frac{\hat{V}'(|\,\boldsymbol{Q}_{n+\alpha}\lambda_e\boldsymbol{e}|)}{|\,\boldsymbol{Q}_{n+\alpha}\lambda_e\boldsymbol{e}|} . \tag{52}$$

From Eq. (51), after some straightforward manipulations, we obtain that the position and momentum on the relative equilibrium must satisfy:

$$\begin{aligned} \boldsymbol{p}_e =& \frac{m}{\Delta t}\left[\boldsymbol{P} - \boldsymbol{I} + \nu(\frac{1}{2}-\beta)\,\boldsymbol{Q}_{n+\alpha} + \nu\beta\boldsymbol{P}\,\boldsymbol{Q}_{n+\alpha}\right]\boldsymbol{q}_e , \\ \boldsymbol{0} =& (\boldsymbol{P}-\boldsymbol{I})^2\boldsymbol{q}_e + \nu(\boldsymbol{P}-\boldsymbol{I})\left[(\frac{1}{2}-\beta)\boldsymbol{I} + \beta\boldsymbol{P}\right]\boldsymbol{Q}_{n+\alpha}\boldsymbol{q}_e \\ & + \nu\,[(1-\gamma)\boldsymbol{I} + \gamma\boldsymbol{P}]\,\boldsymbol{Q}_{n+\alpha}\boldsymbol{q}_e . \end{aligned} \tag{53}$$

To determine if these equations have a solution, let us define the scalars κ_2, κ_1, κ_T, κ_0

$$\begin{aligned} \kappa_0 &= 1+\nu(\frac{1}{2}-2\beta+3\alpha\beta+\gamma-2\alpha\gamma)\;, \\ \kappa_1 &= -2+\nu(-3\alpha\beta+\beta+\frac{\alpha}{2}+\alpha\gamma)\;, \\ \kappa_2 &= 1+\alpha\beta\nu\;, \\ \kappa_T &= \nu\left[(\beta-\frac{1}{2})(1-\alpha)+(1-\gamma)(1-\alpha)\right]\;, \end{aligned} \tag{54}$$

so that, Eq. (53)$_2$ becomes

$$\left[\kappa_2\boldsymbol{P}^2+\kappa_1\boldsymbol{P}+\kappa_T\boldsymbol{P}^T+\kappa_0\boldsymbol{I}\right]\boldsymbol{Q}_n\boldsymbol{q}_e=\boldsymbol{0}\;. \tag{55}$$

Without loss of generality, choose $\boldsymbol{Q}_n\boldsymbol{q}_e=\lambda_e\boldsymbol{e}$, to finally obtain

$$\left[\kappa_2\boldsymbol{P}^2+\kappa_1\boldsymbol{P}+\kappa_T\boldsymbol{P}^T+\kappa_0\boldsymbol{I}\right]\boldsymbol{e}=\boldsymbol{0}\;. \tag{56}$$

The trivial solution $\boldsymbol{P}=\boldsymbol{I}$, $\nu=0$ is a solution for every combination of parameters (α,β,γ), but corresponds to the mass at rest. Any nontrivial solution corresponds to the zero sum of four vectors. See Fig. 5 for an illustration.

In the case of the trapezoidal rule, $(\alpha,\beta,\gamma)=(1,\frac{1}{4},\frac{1}{2})$, the identity (56) evaluates to

$$\begin{aligned} &\kappa_0=\kappa_2=1+\frac{\nu}{4}\;,\quad \kappa_1=-2+\frac{\nu}{2}\;,\quad \kappa_T=0\;, \\ &\boldsymbol{0}=\left[\boldsymbol{P}^2+\frac{\kappa_1}{\kappa_0}\boldsymbol{P}+\boldsymbol{I}\right]\boldsymbol{e}\;. \end{aligned} \tag{57}$$

See Fig. 6 for a graphical interpretation of Eq. (57). Defining $\eta=\kappa_1/\kappa_o$, Eq. (57) can be rewritten as

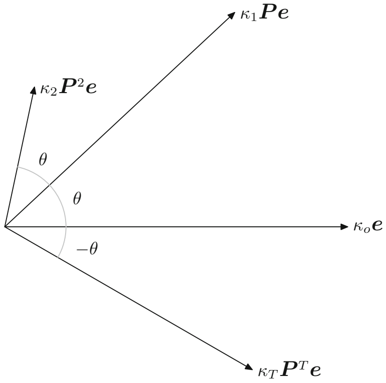

$\kappa_1\boldsymbol{Pe}$
$\kappa_2\boldsymbol{P}^2\boldsymbol{e}$
θ
θ
$\kappa_o\boldsymbol{e}$
$-\theta$
$\kappa_T\boldsymbol{P}^T\boldsymbol{e}$

Fig. 5 Graphical interpretation of Eq. (56). Each term of the equation can be viewed as a vector and their sum must vanish

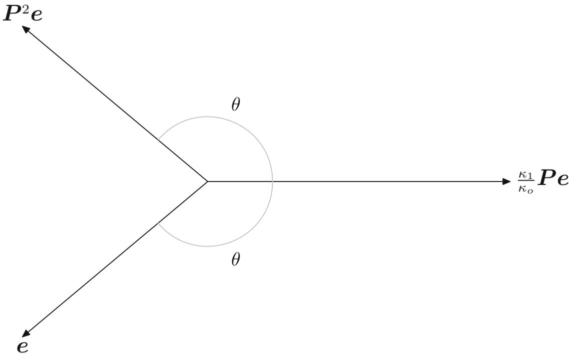


Fig. 6 Graphical interpretation of the relation (57). We identify each term of the equation with a vector, and their sum must vanish

$$\begin{aligned} \cos 2\hat{\theta} + \eta \cos \hat{\theta} &= -1 \ , \\ \sin 2\hat{\theta} + \eta \sin \hat{\theta} &= 0 \ . \end{aligned} \tag{58}$$

These equations have a nontrivial solution $\eta = -2\cos\hat{\theta}$, which implies that the trapezoidal rule possesses solutions which are discrete relative equilibria of the spring-mass problem. For a given $\boldsymbol{q}_e$, the corresponding $\boldsymbol{p}_e$ is recovered from Eq. (53), using the definition of η:

$$\boldsymbol{p}_e = \frac{m}{\Delta t}\begin{bmatrix} 0 & \frac{2\sin\hat{\theta}}{1+\cos\hat{\theta}} \\ -\frac{2\sin\hat{\theta}}{1+\cos\hat{\theta}} & 0 \end{bmatrix} , \qquad \boldsymbol{q}_e = \frac{m}{\Delta t}\sqrt{\nu}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\boldsymbol{q}_e \ . \tag{59}$$

Moreover, the discrete relative equilibria obtained with the trapezoidal rule lie on the exact ones. To see this, observe that by construction, the motion described by Eq. (59) is a discrete orbit of the symmetry group with energy

$$H(z_n) = K(\boldsymbol{p}_n) + V(\boldsymbol{q}_n) = \hat{K}(\pi_e) + \hat{V}(\lambda_e) \ , \tag{60}$$

as the exact relative equilibria. To calculate the corresponding angular momentum, let $\boldsymbol{q}_e = |\boldsymbol{q}_e|\boldsymbol{e}$ and by Eq. (59)

$$\begin{aligned} \boldsymbol{p}_e &= \frac{m}{\Delta t}\sqrt{\nu}|\boldsymbol{q}_e|\boldsymbol{e}_2 \\ \boldsymbol{\mu} = \boldsymbol{q}_e \times \boldsymbol{p}_e &= \frac{m}{\Delta t}\sqrt{\nu}\,|\boldsymbol{q}_e|^2\boldsymbol{e}_3 \end{aligned} \tag{61}$$

implying

$$|\boldsymbol{\mu}|^2 = \frac{m^2}{\Delta t^2}\,\nu\,|\boldsymbol{q}_e|^4 = m\,\hat{V}'(\lambda_e)\,\lambda_e^3 \ , \tag{62}$$

which coincides with the value of the of the angular momentum in the exact relative equilibrium. In summary, the discrete relative equilibria of the trapezoidal rule are discrete group orbits with energy and angular momentum equal to the corresponding values in the exact relative equilibria. Hence, the discrete motions lie on the exact relative equilibria.

The previous analysis is now modified for the parameter combination $(\alpha, \beta, \gamma) = (1, \frac{1}{4}(\frac{1}{2} + \gamma)^2, \gamma)$, with $\frac{1}{2} < \gamma \leq 1$, that corresponds the first-order members of Newmark's method with maximum dissipation. By replacing $\alpha = 1$ in Eq. (54) it follows that

$$\begin{aligned} \kappa_0 &= \quad 1 + \nu(\frac{1}{2} + \beta - \gamma) \ , \\ \kappa_1 &= -2 + \nu(-2\beta + \frac{1}{2} + \gamma) \ , \\ \kappa_2 &= \quad 1 + \nu\beta \\ \kappa_T &= 0 \ . \end{aligned} \tag{63}$$

Inserting these parameters in Eq. (56) gives, after some manipulations,

$$\begin{aligned} \eta_1 &= \frac{-2 + \nu(\frac{1}{2} + \gamma - 2\beta)}{1 + \nu(\beta - \gamma + \frac{1}{2})} \ , \\ \eta_2 &= \frac{(1 + \beta\nu)}{1 + \nu(\beta - \gamma + \frac{1}{2})} \ , \\ & \left[\eta_2 \boldsymbol{P}^2 + \eta_1 \boldsymbol{P} + \boldsymbol{I}\right] \boldsymbol{e} = \boldsymbol{0} \ . \end{aligned} \tag{64}$$

The third equation is illustrated in Fig. 7. From this figure it is clear that a necessary condition for the existence of solutions is that $\eta_2 = 1$. But η_2 is strictly greater than 1 for every $\nu > 0$ and thus it can be concluded that the dissipative family of Newmark's method cannot have solutions which are relative equilibria of the spring-mass problem.

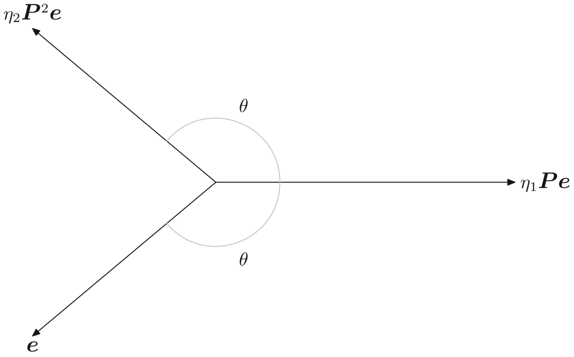

$\eta_2 \boldsymbol{P}^2 \boldsymbol{e}$
θ
$\eta_1 \boldsymbol{P} \boldsymbol{e}$
θ
$\boldsymbol{e}$

Fig. 7 Graphical interpretation of Eq. $(64)_3$. Each of the terms represented by a vector. A necessary condition for their sum to be zero is that the vertical components of $\boldsymbol{e}$ and $\eta_2 \boldsymbol{P}^2 \boldsymbol{e}$ are equal and opposite. This implies $\eta_2 = 1$

To study the HHT method, the same analysis is repeated once more, for algorithmic parameters $(\alpha, \beta, \gamma) = (\alpha, (1-\alpha/2)^2, 3/2-\alpha),\ \ 0.7 \leq \alpha \leq 1$. The value of the scalars $\kappa_0, \kappa_1, \kappa_2, \kappa_T$ is now

$$\begin{aligned}
\kappa_0 &= \ \ \ 1 + \frac{\nu}{4}\alpha(3\alpha^2 - 6\alpha + 4)\ , \\
\kappa_1 &= -2 + \frac{\nu}{4}(-3\alpha^3 + 9\alpha^2 - 8\alpha + 4)\ , \\
\kappa_2 &= \ \ \ 1 + \frac{\nu}{4}\alpha(2-\alpha)^2\ , \\
\kappa_T &= \frac{\nu}{4}\alpha^2(1-\alpha)\ .
\end{aligned} \tag{65}$$

Using these parameters in Eq. (56), this last equation becomes

$$\begin{aligned}
\kappa_2 \cos(2\hat{\theta}) + (\kappa_1 + \kappa_T)\cos\hat{\theta} + \kappa_0 &= 0\ , \\
\kappa_2 \sin(2\hat{\theta}) + (\kappa_1 - \kappa_T)\sin\hat{\theta} \qquad &= 0\ .
\end{aligned} \tag{66}$$

The second equations gives

$$\cos\hat{\theta} = \frac{\kappa_T - \kappa_1}{2\kappa_2} \tag{67}$$

which is plotted in Fig. 8 as a function of α. Inserting Eq. (67) in Eq. $(66)_1$, results in the algebraic equation

$$\kappa_T(\kappa_T - \kappa_1) + \kappa_2(\kappa_0 - \kappa_2) = 0 \Leftrightarrow \frac{1}{4}(\alpha-1)\alpha^2\nu^2 = 0\ . \tag{68}$$

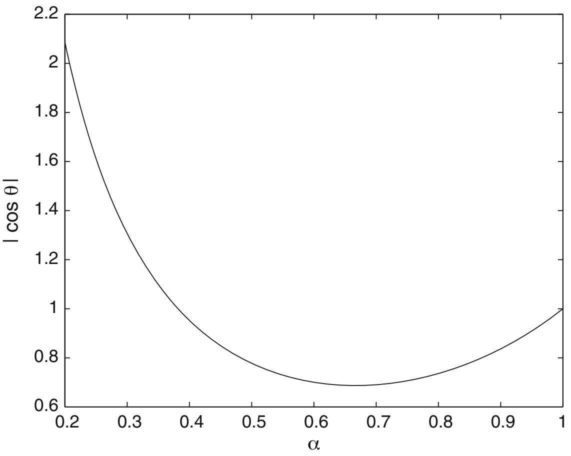


Fig. 8 HHT method. $(\kappa_T - \kappa_1)/(2\kappa_2)$ as a function of α. If this ratio must be equal to $\cos\theta$, the parameter α must belong ot (0.35, 1)

This equation has three solutions: when $\nu = 0$, corresponding to the trivial solution as indicated before, $\alpha = 1$, corresponding to the trapezoidal rule, already studied, and the case $\alpha = 0$. This last solution must be rejected since Fig. 8 shows that for $\alpha < 0.35$ the function $(\kappa_T - \kappa_1)/(2\kappa_2)$ can not be the cosine of any angle. It can be concluded that, as for the dissipative members of Newmark's family, there is no dissipative HHT scheme able to represent a discrete relative equilibrium of the spring-mass system.

6 Summary

This chapter deals with the numerical approximation of the initial boundary value problem of continuum dynamics. More specifically, it presents the most commonly used methods to solve it, i.e. a discretization in space with finite elements followed by an integration in time using methods for ordinary differential equations.

The main emphasis has been the discussion of the formulation and analysis of discretization techniques that employ time integration schemes with controllable, artificial, high frequency dissipation. These are extremely popular in commercial codes for their ability to obtain approximate solutions to complex, stiff, problems in linear and nonlinear mechanics.

The chapter has presented the derivation of space and time discretization, and the most commonly used analysis techniques that can be employed to assess the stability of this type of methods, and their dissipation properties. Since these techniques are different in the linear and nonlinear regimes, we have presented them separately.

In linear problems, spectral analysis has been the *de facto* methodology to analyze time integration schemes. While this is an extremely useful tool, we argue that it might lead to unsupported conclusions both for linear and nonlinear problems. In linear problems, unconditional spectral stability is not sufficient to guarantee unconditional stability of the complete space-time discretization, as it is often believed. An additional property of the method, originally obtained by the author, is reviewed.

Two are the most common methods for analyzing the stability of integration schemes for linear problems, namely, spectral and energy methods. We have presented a third strategy, never used in this context to the author's knowledge, which is based on backward error analysis. Using this method we are able to show that the response of high frequency dissipative schemes closely resembles the exact solution of perturbed systems in which additional, frequency dependent, damping and stiffness are added. This idea sheds new light into the understanding of dissipative time stepping methods.

In nonlinear problems, the only way to assess the stability of a discretization is by evaluating the evolution of the energy in the system. Since spectral analysis can not be used to study this aspect, the focus needs to be shifted toward numerical methods which can guarantee unconditional energy dissipation in the nonlinear regime. One of these types of methods is the EDMC family, which in addition to controlling energy growth for all hyperelastic models, exactly preserve linear and angular momenta in problems with translational and rotational symmetry. The EDMC are recalled in this chapter, and their major conservation properties proven.

A very interesting feature of nonlinear solid and structural dynamics with symmetry is the existence of a particular kind of solutions, known as relative equilibria. These are motions along level sets of constant momenta, energy, and with pointwise constant deformation, stress (and thus strain energy density). We show that EDMC methods preserve these solutions and, moreover, when a trajectory starts close to a relative equilibrium, it will be attracted to it asymptotically. We show, also for the first time in the authors' knowledge, that standard dissipative schemes can not preserve relative equilibria even in the simplest cases, explaining why long term simulations obtained with such type of methods invariantly end up in static equilibria.

Acknowledgments Funding for this work has been provided by the Spanish Ministry of Science and Competitiveness under Grant DPI2012-36429.

References

Armero, F., & Petocz, E. (1999). A new dissipative time-stepping algorithm for frictional contact problems: Formulation and analysis. *Computer Methods in Applied Mechanics and Engineering*, *179*, 151–178.

Armero, F., & Romero, I. (2003). Energy-dissipative momentum-conserving time-stepping algorithms for the dynamics of nonlinear cosserat rods. *Computational Mechanics*, *31*, 3–26.

Armero, F., & Romero, I. (2001a). On the formulation of high-frequency dissipative time-stepping algorithms for nonlinear dynamics. Part I: Low order methods for two model problems and nonlinear elastodynamics. *Computer Methods in Applied Mechanics and Engineering*, *190*, 2603–2649.

Armero, F., & Romero, I. (2001b). On the formulation of high-frequency dissipative time-stepping algorithms for nonlinear dynamics. Part II: Second order methods. *Computer Methods in Applied Mechanics and Engineering*, *190*, 6783–6824.

Bathe, K. J. (1996). *Finite element procedures*. Englewood Cliffs, NJ: Prentice Hall.

Bathe, K. J. (2007). Conserving energy and momentum in nonlinear dynamics: A simple implicit time integration scheme. *Computers and Structures*, *85*(7–8), 437–445.

Bathe, K. J., & Wilson, E. L. (1973). Stability and accuracy analysis of direct integration methods. *Earthquake Engineering and Structural Dynamics*, *1*(1), 283–291.

Bauchau, O. A., & Joo, T. (1999). Computational schemes for non-linear elasto-dynamics. *International Journal for Numerical Methods in Engineering*, *45*(6), 693–719.

Bauchau, O. A., & Theron, N. J. (1996). Energy decaying scheme for non-linear beam models. *Computer Methods in Applied Mechanics and Engineering*, *134*(1–2), 37–56.

Bauchau, O. A., Damilano, G., & Theron, N. J. (1995). Numerical integration of non-linear elastic multi-body systems. *International Journal for Numerical Methods in Engineering*, *38*(16), 2727–2751.

Bazzi, G., & Anderheggen, E. (1982). The ρ-family of algorithms for time-step integration with improved numerical dissipation. *Earthquake Engineering and Structural Dynamics*, *10*, 537–550.

Belytschko, T. (1983). An overview of semidiscretization and time integration procedures. *Computational Methods for Transient Analysis*, 67–155.

Belytschko, T., & Schoeberle, D. F. (1975). On the unconditional stability of an implicit algorithm for nonlinear structural dynamics. *Journal of Applied Mechanics*, *42*, 865–869.

Betsch, P., & Steinmann, P. (2001). Conservation properties of a time FE method. Part II: Time stepping schemes for non-linear elastodynamics. *International Journal for Numerical Methods in Engineering*, *50*, 1931–1955.

Bottasso, C. L., & Borri, M. (1997). Energy preserving/decaying schemes for non-linear beam dynamics using the helicoidal approximation. *Computer Methods in Applied Mechanics and Engineering*, *143*, 393–415.

Chung, J., & Hulbert, G. M. (1993). A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized-α method. *Journal of Applied Mechanics*, *60*, 371–375.

Erlicher, S., Bonaventura, L., & Bursi, O. S. (2002). The analysis of the generalized-α method for non-linear dynamic problems. *Computational Mechanics*, *28*, 83–104.

Geradin, M. (1974). A classification and discussion of integration operators for transient structural response. In *12th Aerospace Sciences Meeting*. Washington, D.C.

Gonzalez, O. (2000). Exact energy-momentum conserving algorithms for general models in non-linear elasticity. *Computer Methods in Applied Mechanics and Engineering*, *190*, 1763–1783.

Gurtin, M. (1981). *An introduction to continuum mechanics*. Academic Press.

Hairer, E. (1994). Backward analysis of numerical integrators and symplectic methods. *Annals of Numerical Mathematics*, *1*, 107–132.

Hairer, E. (1999). Backward error analysis for multistep methods. *Numerische Mathematik*, *84*(2), 199–232.

Hairer, E., & Wanner, G. (1991). *Solving ordinary differential equations II. Stiff and differential-algebraic problems* (1st ed., Vol. 14). Berlin: Springer.

Hilber, H. M., Hughes, T. J. R., & Taylor, R. L. (1977). Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthquake Engineering and Structural Dynamics*, *5*, 283–292.

Hughes, T. J. R. (1976). Stability, convergence and growth and decay of energy of the average acceleration method in nonlinear structural dynamics. *Computers and Structures*, *6*, 313–324.

Hughes, T. J. R. (1983). Analysis of transient algorithms with particular reference to stability behavior. In T. Belytschko & T. J. R. Hughes (Eds.), *Computational methods for transient analysis*, (pp. 67–155). Amsterdam: Elsevier Scientific Publishing Co.

Hughes, T. J. R. (1987). *The finite element method*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Kuhl, D., & Crisfield, M. A. (1999). Energy-conserving and decaying algorithms in non-linear structural dynamics. *International Journal for Numerical Methods in Engineering*, *45*(5), 569–599.

Kuhl, D., & Ramm, E. (1996). Constraint energy momentum algorithm and its application to nonlinear dynamics of shells. *Computer Methods in Applied Mechanics and Engineering*, *136*, 293–315.

Kuhl, D., & Ramm, E. (1999). Generalized Energy-Momentum method for non-linear adaptive shell dynamics. *Computer Methods in Applied Mechanics and Engineering*, *178*, 343–366.

Leimkuhler, B., & Reich, S. (2004). *Simulating Hamiltonian dynamics*. Cambridge University Press.

Marsden, J. E., & Hughes, T. J. R. (1983). *Mathematical foundations of elasticity*. Englewood Cliffs: Prentice-Hall.

Marsden, J. E., & Ratiu, T. S. (1994). *Introduction to Mechanics and Symmetry* (1st ed.). New York: Springer.

Modak, S., & Sotelino, E. D. (2002, July). The generalized method for structural dynamics applications. *Advances in Engineering Software*, *33*(7–10), 565–575.

Newmark, N. M. (1956). A method of computation for structural dynamics. *Journal of the Engineering Mechanics division. ASCE*, *85*, 67–94.

Romero, I. (2002). On the stability and convergence of fully discrete solutions in linear elastodynamics. *Computer Methods in Applied Mechanics and Engineering*, *191*, 3857–3882.

Romero, I. (2004). Stability analysis of linear multistep methods for classical elastodynamics. *Computer Methods in Applied Mechanics and Engineering*, *193*, 2169–2189.

Simo, J. C., & Tarnow, N. (1992). The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics. *Journal of Applied Mathematics and Physics (ZAMP)*, *43*(5), 757–792.

Simo, J. C., & Tarnow, N. (1994). A new energy and momentum conserving algorithm for the nonlinear dynamics of shells. *International Journal for Numerical Methods in Engineering*, *37*(15), 2527–2549.

Simo, J. C., & Wong, K. (1991). Unconditionally stable algorithms for rigid body dynamics that exactly preserve energy and momentum. *International Journal for Numerical Methods in Engineering*, *31*, 19–52.

Simo, J. C., Marsden, J. E., & Krishnaprasad, P. S. (1988). The Hamiltonian structure of nonlinear elasticity: The material and convective representations of solids, rods, and plates. *Archive for Rational Mechanics and Analysis*, *104*(2), 125–183.

Simo, J. C., Posbergh, T. A., & Marsden, J. E. (1991). Stability of relative equilibria. II. Application to nonlinear elasticity. *Archive for Rational Mechanics and Analysis*, *115*(1), 61–100.

Warming, R. F., & Hyett, B. J. (1974). The modified equation approach to stability and accuracy analysis of finite difference methods. *Journal of Computational Physics*, *14*, 159–179.

Wilson, E. L. (1968). A computer program for the dynamic stress analysis of underground structures. Technical Report EERC Report No. 68-1, University of California, Berkeley.

Wood, W. L. (1990). *Practical time-stepping algorithms*. Oxford: Clarendon Press.

Wood, W. L., Bossak, M., & Zienkiewicz, O. C. (1981). An alpha modification of Newmark's method. *International Journal for Numerical Methods in Engineering*, *15*, 1562–1566.

Zhou, X., & Tamma, K. K. (2004). Design, analysis, and synthesis of generalized single step single solve and optimal algorithms for structural dynamics. *International Journal for Numerical Methods in Engineering*, *59*, 597–668.

Zienkiewicz, O. C., & Taylor, R. L. (2005). *The finite element method for solid and structural mechanics* (6th ed.). Oxford, England: Butterworth Heinemann.

Zienkiewicz, O. C., Wood, W. L., Hine, N. W., & Taylor, R. L. (1984). A unified set of single step algorithms. Part 1: General formulation and applications. *International Journal for Numerical Methods in Engineering*, *20*, 1529–1552.

Energy-Momentum Integrators for Elastic Cosserat Points, Rigid Bodies, and Multibody Systems

Peter Betsch

Abstract The goal of this chapter is to present the development of energy-momentum (EM) schemes in the framework of discrete (or finite-dimensional) mechanical systems. EM integrators belong to the class of structure-preserving numerical methods and have been originally developed in the field of nonlinear solid and structural mechanics. EM schemes and energy dissipating variants thereof typically exhibit improved numerical stability and robustness when compared to standard integrators. Due to their superior numerical properties, EM schemes have soon been extended to more involved applications such as flexible multibody dynamics and coupled thermomechanical problems. In this chapter, we start the development of second-order EM schemes in the context of the Cosserat point (or pseudo-rigid body). The theory of a Cosserat point shares main structural properties with semi-discrete formulations of elastodynamics. Indeed, the Cosserat point can be directly linked to the 4-node tetrahedral finite element. Besides its usefulness in explaining main ingredients of EM schemes such as the algorithmic stress formula, the Cosserat point is ideally suited to perform the transition to rigid body dynamics. In particular, in the present work, the rigid body formulation is obtained by imposing the zero strain condition on the Cosserat point. This way the rigid body is treated as constrained mechanical system. Moreover, we show that the EM discretization of constrained mechanical systems can be derived in a straightforward way from the EM scheme for the Cosserat point. The resulting rigid body formulation is closely connected to natural coordinates. Eventually, we deal with the extension to multibody systems which can be done in a straightforward way due to the presence of holonomic constraints in the present rigid body formulation.

Funding for this work has been provided by the German Science Foundation under Grant BE 2285/10-1.

P. Betsch (✉)
Institute of Mechanics, Karlsruhe Insitute of Technology, Karlsruhe, Germany
e-mail: peter.betsch@kit.edu

P. Betsch (ed.), *Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics*, CISM International Centre for Mechanical Sciences 565, DOI 10.1007/978-3-319-31879-0_2

1 Introduction

Energy-momentum (EM) integrators have been originally developed in the context of nonlinear structural dynamics. Building upon the previous work by Hughes et al. (1978), Greenspan (1984), Simo and Wong (1991) and Simo et al. (1992b), the first EM scheme for nonlinear elastodynamics has been proposed in the seminal work by Simo and Tarnow (1992). Due to their favorable numerical stability properties (Gonzalez and Simo 1996) EM methods have soon been extended to the realm of nonlinear structural and rigid body dynamics. The description of these systems typically relies on the introduction of rotational coordinates. For example, in rigid body dynamics one may use Euler angles, Euler parameters (or unit quaternions), or Rodrigues parameters for the parametrization of the rotation manifold. It was soon realized that the selection of specific rotational coordinates has a strong impact on the design of structure-preserving integrators (Lewis and Simo 1994). In particular, the use of minimal coordinates (like Euler angles for rigid body dynamics) in general leads to highly nonlinear and elaborate expressions that typically impede the design of time-stepping schemes featuring conservation of angular momentum.

In nonlinear structural dynamics the parametrization of the rotation manifold does affect both the discretization in space and time. It has been shown in Simo et al. (1992a) that the finite element interpolation of rotational variables in general destroys conservation of angular momentum of the semi-discrete system. Strictly speaking the available EM methods for nonlinear beams (Romero and Armero 2002a; Betsch and Steinmann 2003; Leyendecker et al. 2006) and shells (Simo and Tarnow 1994; Brank et al. 1998; Betsch and Sänger 2009a) confine the use of rotational parameters to the nodes of the finite element mesh. Similarly, in these works the discretization in time does not directly rely on the use of rotational parameters.

EM schemes provide a good starting point for the development of energy decaying schemes. Energy decaying variants of EM schemes have been proposed, for example, by Bauchau and Bottasso (1999), Kuhl and Crisfield (1999), Armero and Romero (2001, 2003), Romero and Armero (2002b), Bottasso et al. (2002), Bottasso and Trainelli (2004), Lens and Cardona (2007). More about energy decaying integrators can be found in chapters "High Frequency Dissipative Integration Schemes for Linear and Nonlinear Elastodynamics" and "A Lie Algebra Approach to Lie Group Time Integration of Constrained Systems".

Due to their desirable numerical properties, EM methods have also been extended to more involved problems such as nonlinear visco-elastodynamics (Groß and Betsch 2010), thermo-elastodynamics (Romero 2009; Groß and Betsch 2011; Romero 2010; Hesch and Betsch 2011c; Conde Martín et al. 2016), finite deformation contact problems (Laursen and Chawla 1997; Armero and Petöcz 1998; Hesch and Betsch 2009, 2011b), and flexible multibody dynamics (Bauchau and Bottasso 1999; Ibrahimbegović et al. 2000; Bottasso et al. 2001; Betsch and Steinmann 2002a, c; Lens et al. 2004; Betsch and Sänger 2009b; Leyendecker et al. 2008a). EM schemes have also been incorporated into direct methods for the optimal control of multibody systems (Bottasso and Croce 2004; Betsch et al. 2012; Koch and Leyendecker 2013).

A good account on the development of EM schemes in the context of nonlinear finite element methods can be found in the books by Crisfield (1997), Géradin and Cardona (2001), Laursen (2002), Krenk (2009), Ibrahimbegović (2009), Bauchau (2011).

An alternative route to the design of structure-preserving time-stepping schemes are variational integrators. In the context of multibody dynamics variational integrators have been dealt with, for example, in Leyendecker et al. (2008b), Ober-Blöbaum et al. (2011), Leyendecker et al. (2010), Johnson and Murphey (2009), Betsch et al. (2010). For a tutorial on variational integrators we refer to the chapter "A Brief Introduction to Variational Integrators".

The goal of this chapter is to present the development of EM schemes in the framework of discrete (or finite-dimensional) mechanical systems. To this end, we start in Sect. 2 with the Cosserat point (or pseudo-rigid body). The theory of a Cosserat point shares main structural properties with semi-discrete formulations of elastodynamics. Indeed, the Cosserat point can be directly linked to the 4-node tetrahedral finite element as will be shown in Sect. 2.6. Besides its usefulness in explaining main ingredients of EM schemes such as the algorithmic stress formula, the Cosserat point is ideally suited to perform the transition to rigid body dynamics. In Sect. 3, the rigid body formulation is obtained by imposing the zero strain condition on the Cosserat point. This way the rigid body is treated as constrained mechanical system. Moreover, the EM discretization of constrained mechanical systems can be derived in a straightforward way from the previously developed EM scheme for the Cosserat point. The resulting rigid body formulation is closely connected to natural coordinates as will be shown in Sect. 3.7. Due to the presence of holonomic constraints in the present rigid body formulation, the extension to multibody systems can be done in a straightforward way. This is the subject of Sect. 4. Eventually, in Sect. 5, representative numerical examples are presented.

2 EM Method for Cosserat Points

We start the description of EM schemes in the context of a Cosserat point (Rubin 2000). Similar to the theory of a pseudo-rigid body (Cohen and Muncaster 1988; Nordenholz and O'Reilly 1998) the theory of a Cosserat point represents a finite-dimensional model for a deformable body. This model problem already features key structural properties of more complicated mechanical systems such as nonlinear elastodynamics and structural dynamics. In contrast to the continuum theory the equations governing the motion of a Cosserat point consist of ordinary differential equations (ODEs). Due to its relative simplicity, the theory of a Cosserat point is deemed to be especially well-suited to convey main ideas of the design of EM schemes.

In addition to that, the theory of a Cosserat point paves the way to rigid body dynamics. To this end additional geometric constraints are imposed on the Cosserat point leading to differential-algebraic equations (DAEs) governing the motion of a rigid body. Consequently, we shall regard the rigid body as a constrained mechanical system. The DAEs not only govern the motion of rigid bodies but also the motion of general flexible multibody systems. It will subsequently become apparent that the development of EM methods for constrained mechanical systems is closely related to the design of EM schemes for elastic bodies.

2.1 Governing Equations

The equations of motion pertaining to the present model problem of a deformable body can be derived from the principle of virtual work for a general deformable continuum. To this end, the assumption of spatially homogeneous deformations is imposed by considering affine deformation maps of the form (Fig. 1)

$$\boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{X}, t) = \overline{\boldsymbol{x}}(t) + \boldsymbol{F}(t)(\boldsymbol{X} - \overline{\boldsymbol{X}}) \tag{1}$$

Here, material points in the reference configuration $\mathcal{B} \subset \mathbb{R}^3$ are denoted by $\boldsymbol{X} \in \mathcal{B}$, $\overline{\boldsymbol{X}} \in \mathcal{B}$ is the center of mass, and $\overline{\boldsymbol{x}}(t) \in \mathcal{B}_t$ denotes the corresponding placement in the configuration $\mathcal{B}_t \subset \mathbb{R}^3$ at time $t \in [0, T]$, the time interval of interest. Moreover, $\boldsymbol{F}(t) = D\boldsymbol{\Phi}_t(\boldsymbol{X})$ is the deformation gradient, where $\boldsymbol{\Phi}_t(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X}, t)$.

Due to the kinematic assumption (1) the deformation gradient $\boldsymbol{F}$ does not depend on $\boldsymbol{X}$. This property gives rise to the present model problem of a Cosserat point. Alternatively, the model problem could be termed *homogeneous elasticity* (see, for example, Simo et al. 1991). Another perspective is to view the present model problem as an extension of the classical model of a rigid body. This viewpoint leads to the notion of a pseudo-rigid body. The configuration space of the free Cosserat point is given by

$$\mathsf{Q} = \left\{(\overline{\boldsymbol{x}}, \boldsymbol{F}) \in \mathbb{R}^3 \times \mathbb{R}^{3\times 3} \mid \det(\boldsymbol{F}) > 0\right\} \tag{2}$$

Note that $\boldsymbol{F} \in GL^+(3)$, where $GL^+(3)$ is the subgroup of the general linear group, $GL(3)$, consisting of 3×3 matrices with positive determinant. Obviously, $\dim(\mathsf{Q}) = 12$, so that the free Cosserat point has $n = 12$ degrees of freedom (DOFs). We further remark that consistent with the definition of the center of mass in the reference configuration we have the relationships

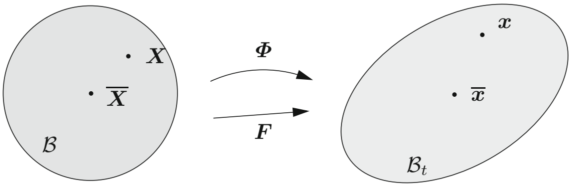


Fig. 1 Planar illustration of the Cosserat point: Reference configuration $\mathcal{B}$ (*left*), and current configuration $\mathcal{B}_t$ (*right*)

$$\overline{\boldsymbol{X}} = \frac{1}{M} \int_{\mathcal{B}_0} \rho_0 \boldsymbol{X} \, dV \qquad \text{or} \qquad \int_{\mathcal{B}_0} \rho_0 (\boldsymbol{X} - \overline{\boldsymbol{X}}) \, dV = \boldsymbol{0}$$

where $M = \int_{\mathcal{B}_0} \rho_0 \, dV$ is the total mass and $\rho_0 : \mathcal{B}_0 \to \mathbb{R}$ is the reference density. The principle of virtual work for a general continuum body can be written as

$$G(\boldsymbol{\Phi}; \delta\boldsymbol{\Phi}) = G_{\text{dyn}}(\boldsymbol{\Phi}; \delta\boldsymbol{\Phi}) + G_{\text{int}}(\boldsymbol{\Phi}; \delta\boldsymbol{\Phi}) - G_{\text{ext}}(\boldsymbol{\Phi}; \delta\boldsymbol{\Phi}) = 0 \tag{3}$$

where $\delta\boldsymbol{\Phi} : \mathcal{B}_0 \to \mathbb{R}^3$ can be interpreted as virtual displacement of the material point $\boldsymbol{X}$, G_{ext} is the virtual work of the external loading, G_{int} is the internal virtual work due to deformation, and G_{dyn} is the contribution of the inertia terms. In particular

$$G_{\text{dyn}}(\boldsymbol{\Phi}; \delta\boldsymbol{\Phi}) = \int_{\mathcal{B}_0} \rho_0 \delta\boldsymbol{\Phi} \cdot \ddot{\boldsymbol{\Phi}} \, dV \tag{4}$$

where $\ddot{\boldsymbol{\Phi}} = \frac{\partial^2}{\partial t^2} \boldsymbol{\Phi}(\boldsymbol{X}, t)$ is the acceleration of the material point $\boldsymbol{X}$ at time t. The virtual work of the internal forces is given by

$$G_{\text{int}}(\boldsymbol{\Phi}; \delta\boldsymbol{\Phi}) = \int_{\mathcal{B}_0} \delta\boldsymbol{F} : \boldsymbol{P} \, dV \tag{5}$$

where $\delta\boldsymbol{F} = D\delta\boldsymbol{\Phi}(\boldsymbol{X})$, and $\boldsymbol{P}$ is the first Piola–Kirchhoff stress tensor. Note that $\boldsymbol{P} = \boldsymbol{F}\boldsymbol{S}$, where $\boldsymbol{S}$ is the second Piola–Kirchhoff stress tensor. We further remark that the scalar product of the two second-order tensors $\delta\boldsymbol{F}$ and $\boldsymbol{P}$ is given by

$$\delta\boldsymbol{F} : \boldsymbol{P} = \text{tr}(\delta\boldsymbol{F}^T \boldsymbol{P})$$

where tr(•) is the trace operator and $\delta\boldsymbol{F}^T$ denotes the transpose of $\delta\boldsymbol{F}$. The virtual work of the external loading can be written as

$$G_{\text{ext}}(\boldsymbol{\Phi}; \delta\boldsymbol{\Phi}) = \int_{\mathcal{B}_0} \rho_0 \delta\boldsymbol{\Phi} \cdot \boldsymbol{b} \, dV + \int_{\partial\mathcal{B}_0} \delta\boldsymbol{\Phi} \cdot \boldsymbol{p} \, dA \tag{6}$$

where $\boldsymbol{b} : \mathcal{B}_0 \times [0, T] \to \mathbb{R}^3$ is the body force per unit mass and $\boldsymbol{p} : \partial\mathcal{B}_0 \times [0, T] \to \mathbb{R}^3$ is the nominal traction vector on the boundary. For simplicity of exposition we confine our attention to the pure Neumann problem (i.e., no Dirichlet boundary conditions).

To derive the variational formulation of the present model problem we insert (1) along with

$$\delta\boldsymbol{\Phi}(\boldsymbol{X}) = \delta\overline{\boldsymbol{x}} + \delta\boldsymbol{F}(\boldsymbol{X} - \overline{\boldsymbol{X}}) \tag{7}$$

into the principle of virtual work (3). Accordingly, the virtual work of the inertia terms (4) yields

$$\hat{G}_{\text{dyn}}\big((\overline{\boldsymbol{x}}, \boldsymbol{F}); (\delta\overline{\boldsymbol{x}}, \delta\boldsymbol{F})\big) = \delta\overline{\boldsymbol{x}} \cdot M\ddot{\overline{\boldsymbol{x}}} + \delta\boldsymbol{F} : (\ddot{\boldsymbol{F}}\boldsymbol{E}_0) \tag{8}$$

where $\boldsymbol{E}_0$ is the constant and positive-definite tensor given by

$$\boldsymbol{E}_0 = \int_{\mathcal{B}_0} \rho_0 (\boldsymbol{X} - \overline{\boldsymbol{X}}) \otimes (\boldsymbol{X} - \overline{\boldsymbol{X}})\, dV \tag{9}$$

Note that $\otimes$ is the standard tensor product of two vectors. Tensor (9) is often called the (referential) *Euler tensor* (Gurtin 1981), and is closely related to the classical inertia tensor of rigid body dynamics, see Sect. 3.4 for further details.

Concerning the internal virtual work (5) the assumption of an homogeneous deformation leads to the expression

$$\hat{G}_{\text{int}}(\boldsymbol{F}; \delta\boldsymbol{F}) = \delta\boldsymbol{F} : \left(\boldsymbol{F} \int_{\mathcal{B}_0} \boldsymbol{S}(\boldsymbol{F})\, dV \right) \tag{10}$$

where an elastic solid with stress response function $\boldsymbol{S}(\boldsymbol{F})$ has been assumed. In the following we focus on constitutive models for hyperelastic solids. In particular, a frame-indifferent hyperelastic stress response is given by

$$\boldsymbol{S}(\boldsymbol{F}) = 2DW(\boldsymbol{C}) \tag{11}$$

where W denotes the strain energy density and $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}$ is the right Cauchy–Green deformation tensor. Expression (10) shows that only the stress resultants

$$\overline{\boldsymbol{S}} = \int_{\mathcal{B}_0} \boldsymbol{S}\, dV = 2V_0 DW(\boldsymbol{C}) \tag{12}$$

enter the internal virtual work. Here $V_0 = \int_{\mathcal{B}_0} dV$ is the total volume of the body in the reference configuration. We further introduce the total strain energy given by

$$U = \int_{\mathcal{B}_0} W(\boldsymbol{C})\, dV = V_0 W(\boldsymbol{C}) \tag{13}$$

so that the second Piola–Kirchhoff stress resultants (12) can be written as

$$\overline{\boldsymbol{S}} = 2DU(\boldsymbol{C}) \tag{14}$$

Now the internal virtual work pertaining to the hyperelastic Cosserat point can be written as

$$\hat{G}_{\text{int}}(\boldsymbol{F}; \delta\boldsymbol{F}) = \delta\boldsymbol{F} : \big(2\boldsymbol{F}DU(\boldsymbol{C})\big) \tag{15}$$

The virtual work of the external loading (6) together with (7) yields

$$\hat{G}_{\text{ext}}\big((\overline{\boldsymbol{x}}, \boldsymbol{F}); (\delta\overline{\boldsymbol{x}}, \delta\boldsymbol{F})\big) = \delta\overline{\boldsymbol{x}} \cdot \boldsymbol{f}_{\text{ext}} + \delta\boldsymbol{F} : \boldsymbol{M}_{\text{ext}} \tag{16}$$

where

$$\boldsymbol{f}_{\text{ext}} = \int_{\mathcal{B}_0} \rho_0 \boldsymbol{b} \, dV + \int_{\partial\mathcal{B}_0} \boldsymbol{p} \, dA \tag{17}$$

$$\boldsymbol{M}_{\text{ext}} = \int_{\mathcal{B}_0} \rho_0 \boldsymbol{b} \otimes (\boldsymbol{X} - \overline{\boldsymbol{X}}) \, dV + \int_{\partial\mathcal{B}_0} \boldsymbol{p} \otimes (\boldsymbol{X} - \overline{\boldsymbol{X}}) \, dA \tag{18}$$

Note that $\boldsymbol{f}_{\text{ext}}$ is the resultant external force acting on the body. Moreover, $\boldsymbol{M}_{\text{ext}}$ may be called *referential external force-moment* (Cohen and Muncaster 1988) relative to the center of mass. Altogether the variational formulation emanating from the principle of virtual work (3) can be written as

$$\delta\overline{\boldsymbol{x}} \cdot \left(M\ddot{\overline{\boldsymbol{x}}} - \boldsymbol{f}_{\text{ext}}\right) + \delta\boldsymbol{F} : \left(\ddot{\boldsymbol{F}}\boldsymbol{E}_0 + 2\boldsymbol{F}DU(\boldsymbol{C}) - \boldsymbol{M}_{\text{ext}}\right) = 0 \tag{19}$$

The last equation has to hold for arbitrary $\delta\overline{\boldsymbol{x}} \in \mathbb{R}^3$ and $\delta\boldsymbol{F} \in \mathbb{R}^{3\times3}$. These equations give rise to the following initial value problem: Find $\overline{\boldsymbol{x}} : [0, T] \to \mathbb{R}^3$ and $\boldsymbol{F} : [0, T] \to \mathbb{R}^{3\times3}$ such that

$$\begin{aligned} M\ddot{\overline{\boldsymbol{x}}} &= \boldsymbol{f}_{\text{ext}} \\ \ddot{\boldsymbol{F}}\boldsymbol{E}_0 + 2\boldsymbol{F}DU(\boldsymbol{C}) &= \boldsymbol{M}_{\text{ext}} \end{aligned} \tag{20}$$

subject to the initial conditions $\overline{\boldsymbol{x}}(0) = \overline{\boldsymbol{x}}_0$, $\dot{\overline{\boldsymbol{x}}}(0) = \overline{\boldsymbol{v}}_0$, $\boldsymbol{F}(0) = \boldsymbol{F}_0$, and $\dot{\boldsymbol{F}}(0) = \boldsymbol{V}_0$, where $\overline{\boldsymbol{x}}_0, \overline{\boldsymbol{v}}_0 \in \mathbb{R}^3$ and $\boldsymbol{F}_0, \boldsymbol{V}_0 \in \mathbb{R}^{3\times3}$ are given quantities. The above ODEs coincide with the *equations of motion in referential form* in Cohen and Muncaster (1988).

2.1.1 Balance Laws

Before dealing with the balance laws we recast (19) in an alternative form. Note, however, that the balance laws for linear momentum, angular momentum, and energy can be directly deduced from the principle of virtual work in the form (19) as well. For completeness, this procedure is outlined in Appendix A.1.

2.2 Formulation in Terms of Directors

For our purposes it is convenient to recast the previously derived equations of motion in a form that is typically used in the theory of a Cosserat point (Rubin 2000). To this end we write the homogeneous deformation gradient as

$$\boldsymbol{F}(t) = \boldsymbol{d}_i(t) \otimes \boldsymbol{D}^i \tag{21}$$

where $\boldsymbol{d}_i(t) \in \mathbb{R}^3$ are three director vectors that in general rotate, stretch and shear with the body in its motion (Fig. 2). Note that in the last equation and in what follows, the summation convention applies to indices appearing twice in a formula. The directors are subject to the requirement that $(\boldsymbol{d}_1 \times \boldsymbol{d}_2) \cdot \boldsymbol{d}_3 > 0$, which is consistent with the condition $\det(\boldsymbol{F}) > 0$.

Corresponding to the directors $\boldsymbol{d}_i(t) \in \mathbb{R}^3$, we introduce the vectors $\boldsymbol{D}_i \in \mathbb{R}^3$ constituting the director triad in the reference configuration. In particular, $\boldsymbol{D}_i$ represent a basis fixed in space whose origin coincides with the center of mass. The corresponding coordinates will be denoted by

$$X^i = \boldsymbol{D}^i \cdot (\boldsymbol{X} - \overline{\boldsymbol{X}}) \tag{22}$$

Without loss of generality, we assume that the directors in the reference configuration are mutually orthonormal, that is, $\boldsymbol{D}^i \cdot \boldsymbol{D}^j = \delta^{ij}$, where δ^{ij} is the Kronecker delta. Note that the last assumption implies $\boldsymbol{D}_i = \boldsymbol{D}^i$. We further assume that the reference configuration of the Cosserat point is a natural (i.e., stress-free) configuration. Inserting (21) into (1) and taking into account (22) yields

$$\boldsymbol{x} = \overline{\boldsymbol{x}}(t) + X^i \boldsymbol{d}_i(t) \tag{23}$$

The last equation indicates that the kinematics of the Cosserat point confines the (convected) coordinates X^i to remain straight. Differentiating (21) with respect to time gives

$$\begin{aligned} \dot{\boldsymbol{F}}(t) &= \dot{\boldsymbol{d}}_i(t) \otimes \boldsymbol{D}^i \\ \ddot{\boldsymbol{F}}(t) &= \ddot{\boldsymbol{d}}_i(t) \otimes \boldsymbol{D}^i \end{aligned} \tag{24}$$

In line with (7) we further have

$$\delta\boldsymbol{F} = \delta\boldsymbol{d}_i \otimes \boldsymbol{D}^i \tag{25}$$

Now we are in a position to recast the ODEs (20) governing the motion of the Cosserat point in an alternative form. Substituting (25) along with $(24)_2$ into the virtual work (8) of the inertia terms we obtain

$$\tilde{G}_{\text{dyn}}\big((\overline{\boldsymbol{x}}, \boldsymbol{d}_i); (\delta\overline{\boldsymbol{x}}, \delta\boldsymbol{d}_i)\big) = \delta\overline{\boldsymbol{x}} \cdot M\ddot{\overline{\boldsymbol{x}}} + \delta\boldsymbol{d}_i \cdot E_0^{ij}\ddot{\boldsymbol{d}}_j \tag{26}$$

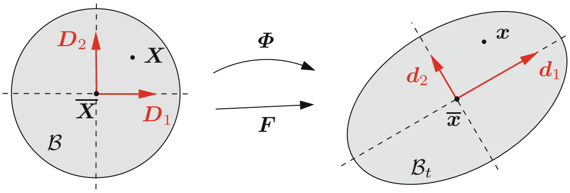


Fig. 2 Planar illustration of the Cosserat point: Reference configuration $\mathcal{B}$ (*left*), and current configuration $\mathcal{B}_t$ (*right*)

where the components E_0^{ij} of the referential Euler tensor (9) are given by

$$E_0^{ij} = \boldsymbol{D}^i \cdot \boldsymbol{E}_0 \boldsymbol{D}^j = \int_{\mathcal{B}_0} \rho_0 X^i X^j \, dV \tag{27}$$

In the last equation use has been made of (9) and (22). Similarly, expression (15) for the internal virtual work can be recast in the form

$$\tilde{G}_{\text{int}}(\boldsymbol{d}; \delta \boldsymbol{d}_i) = \delta \boldsymbol{d}_i \cdot \boldsymbol{d}_j \, (\boldsymbol{D}^i \cdot 2DU(\boldsymbol{C})\boldsymbol{D}^j) \tag{28}$$

where use has been made of (13). Note that the strain energy $U(\boldsymbol{C})$ depends on the right Cauchy–Green deformation tensor $\boldsymbol{C} = \boldsymbol{F}^T \boldsymbol{F}$ which, in view of (21), can also be written as

$$\boldsymbol{C} = d_{ij} \boldsymbol{D}^i \otimes \boldsymbol{D}^j$$

where

$$d_{ij} = \boldsymbol{d}_i \cdot \boldsymbol{d}_j \tag{29}$$

play the role of metric coefficients. Accordingly, we have six independent metric coefficients measuring homogeneous deformations of the Cosserat point. Specifically, if the magnitude of $\boldsymbol{d}_i$ changes, the Cosserat point experiences extension, whereas shear deformation happens if the angle between any two directors changes. Next consider

$$\tfrac{d}{dt} U(\boldsymbol{C}) = DU(\boldsymbol{C}) : \dot{\boldsymbol{C}} = DU(\boldsymbol{C}) : (\dot{d}_{ij} \boldsymbol{D}^i \otimes \boldsymbol{D}^j) = \tfrac{1}{2} \overline{S}^{ij} \dot{d}_{ij} \tag{30}$$

where the components

$$\overline{S}^{ij} = \boldsymbol{D}^i \cdot 2DU(\boldsymbol{C}) \boldsymbol{D}^i = 2 \frac{\partial U}{\partial d_{ij}} \tag{31}$$

have been introduced. Moreover, employing (29) in (30) yields the relationship

$$\tfrac{d}{dt} U(\boldsymbol{C}) = \tfrac{1}{2} \overline{S}^{ij} (\dot{\boldsymbol{d}}_i \cdot \boldsymbol{d}_j + \boldsymbol{d}_i \cdot \dot{\boldsymbol{d}}_j) = \overline{S}^{ij} \boldsymbol{d}_j \cdot \dot{\boldsymbol{d}}_i \tag{32}$$

where the symmetry of $\overline{S}^{ij}$ (i.e., $\overline{S}^{ij} = \overline{S}^{ji}$) has been taken into account. Next we introduce the internal director forces

$$\boldsymbol{f}^i_{\text{int}} = \overline{S}^{ij} \boldsymbol{d}_j = \frac{\partial U}{\partial \boldsymbol{d}_i} \tag{33}$$

such that the internal virtual work (28) can be written as

$$\tilde{G}_{\text{int}}(\boldsymbol{d}; \delta \boldsymbol{d}_i) = \delta \boldsymbol{d}_i \cdot \boldsymbol{f}^i_{\text{int}} \tag{34}$$

Remark 2.1 For later use we note that applying the chain rule and taking into account the symmetry of the metric coefficients d_{jk} one gets

$$\begin{aligned}\frac{\partial U}{\partial \boldsymbol{d}_i} &= \frac{\partial U}{\partial d_{jk}} \frac{\partial d_{jk}}{\partial \boldsymbol{d}_i} \\ &= \frac{\partial U}{\partial d_{jk}} \left(\delta^i_j \boldsymbol{d}_k + \delta^i_k \boldsymbol{d}_j \right) \\ &= 2 \frac{\partial U}{\partial d_{ik}} \boldsymbol{d}_k\end{aligned} \tag{35}$$

This result is consistent with (31) and (33).

The virtual work of the external loading (16) can be written as

$$\tilde{G}_{\mathrm{ext}}\big((\overline{\boldsymbol{x}}, \boldsymbol{d}_i); (\delta\overline{\boldsymbol{x}}, \delta\boldsymbol{d}_i)\big) = \delta\overline{\boldsymbol{x}} \cdot \boldsymbol{f}_{\mathrm{ext}} + \delta\boldsymbol{d}_i \cdot \boldsymbol{f}^i_{\mathrm{ext}} \tag{36}$$

where the external director forces $\boldsymbol{f}^i_{\mathrm{ext}} \in \mathbb{R}^3$ are given by

$$\boldsymbol{f}^i_{\mathrm{ext}} = \boldsymbol{M}_{\mathrm{ext}} \boldsymbol{D}^i = \int_{\mathcal{B}_0} \rho_0 X^i \boldsymbol{b}\, dV + \int_{\partial\mathcal{B}_0} X^i \boldsymbol{p}\, dA \tag{37}$$

To get the last equation use has been made of (18) along with (22). Note that with regard to the last equation the referential external force-moment $\boldsymbol{M}_{\mathrm{ext}}$ defined in (18) can also be written as

$$\boldsymbol{M}_{\mathrm{ext}} = \boldsymbol{f}^i_{\mathrm{ext}} \otimes \boldsymbol{D}_i \tag{38}$$

Now, using (26), (34), and (36), a variational formulation of the Cosserat point equivalent to (19) can be obtained:

$$\delta\overline{\boldsymbol{x}} \cdot \left(M \ddot{\overline{\boldsymbol{x}}} - \boldsymbol{f}_{\mathrm{ext}} \right) + \delta\boldsymbol{d}_i \cdot \left(E^{ij}_0 \ddot{\boldsymbol{d}}_j + \boldsymbol{f}^i_{\mathrm{int}} - \boldsymbol{f}^i_{\mathrm{ext}} \right) = 0 \tag{39}$$

Due to the arbitrariness of $\delta\overline{\boldsymbol{x}} \in \mathbb{R}^3$ and $\delta\boldsymbol{d}_i \in \mathbb{R}^3$, $i = 1, 2, 3$, we obtain 12 independent ODEs giving rise to the following initial value problem for the hyperelastic Cosserat point: Find $\overline{\boldsymbol{x}} : [0, T] \to \mathbb{R}^3$ and $\boldsymbol{d}_i : [0, T] \to \mathbb{R}^3$ $(i = 1, 2, 3)$ such that

$$\begin{aligned} M \ddot{\overline{\boldsymbol{x}}} &= \boldsymbol{f}_{\mathrm{ext}} \\ E^{ij}_0 \ddot{\boldsymbol{d}}_j + \boldsymbol{f}^i_{\mathrm{int}} &= \boldsymbol{f}^i_{\mathrm{ext}} \end{aligned} \tag{40}$$

subject to the initial conditions $\overline{\boldsymbol{x}}(0) = \overline{\boldsymbol{x}}_0$, $\dot{\overline{\boldsymbol{x}}}(0) = \overline{\boldsymbol{v}}_0$, $\boldsymbol{d}_i(0) = (\boldsymbol{d}_i)_0$, and $\dot{\boldsymbol{d}}_i(0) = (\boldsymbol{v}_i)_0$, where $\overline{\boldsymbol{x}}_0$, $\overline{\boldsymbol{v}}_0$, $(\boldsymbol{d}_i)_0$, $(\boldsymbol{v}_i)_0 \in \mathbb{R}^3$ are given quantities. It is worth noting that the ODEs $(40)_2$ coincide with the *balances of director momentum* in Rubin (2000). Moreover, in Rubin (2000), $\boldsymbol{f}^i_{\mathrm{ext}} \in \mathbb{R}^3$ and $\boldsymbol{f}^i_{\mathrm{int}} \in \mathbb{R}^3$ are called *external director couples* and *intrinsic director couples*, respectively.

2.3 Balance of Linear Momentum

The balance law for linear momentum can be directly obtained from the principle of virtual work (39) by setting $\delta\bar{\boldsymbol{x}} = \boldsymbol{\xi}$, where $\boldsymbol{\xi} \in \mathbb{R}^3$ is a constant vector, together with $\delta\boldsymbol{d}_i = \boldsymbol{0}$. Accordingly, we get

$$\frac{d}{dt}\boldsymbol{l} = \boldsymbol{f}_{\text{ext}} \tag{41}$$

where the total linear momentum of the Cosserat point is given by

$$\boldsymbol{l} = M\dot{\bar{\boldsymbol{x}}} \tag{42}$$

As before, the right-hand side of (41) characterizes the resultant external force applied to the Cosserat point.

2.4 Balance of Angular Momentum

In preparation for the design of EM integrators, we next consider the fundamental balance law for angular momentum. Substituting $\delta\bar{\boldsymbol{x}} = \boldsymbol{\xi} \times \bar{\boldsymbol{x}}$ along with $\delta\boldsymbol{d}_i = \boldsymbol{\xi} \times \boldsymbol{d}_i$ into (39) yields

$$(\boldsymbol{\xi} \times \bar{\boldsymbol{x}}) \cdot \left(M\ddot{\bar{\boldsymbol{x}}} - \boldsymbol{f}_{\text{ext}}\right) + (\boldsymbol{\xi} \times \boldsymbol{d}_i) \cdot \left(E_0^{ij}\ddot{\boldsymbol{d}}_j + \boldsymbol{f}^i_{\text{int}} - \boldsymbol{f}^i_{\text{ext}}\right) = 0 \tag{43}$$

or

$$\boldsymbol{\xi} \cdot \left(M\bar{\boldsymbol{x}} \times \ddot{\bar{\boldsymbol{x}}} - \bar{\boldsymbol{x}} \times \boldsymbol{f}_{\text{ext}} + E_0^{ij}\boldsymbol{d}_i \times \ddot{\boldsymbol{d}}_j + \overline{S}^{ij}\boldsymbol{d}_i \times \boldsymbol{d}_j - \boldsymbol{d}_i \times \boldsymbol{f}^i_{\text{ext}}\right) = 0 \tag{44}$$

In the last equation use has been made of (33). Due to the symmetry of $\overline{S}^{ij}$ and the skew-symmetry of the vector cross product we have $\overline{S}^{ij}\boldsymbol{d}_i \times \boldsymbol{d}_j = \boldsymbol{0}$. Now equation (44) can be recast in the form

$$\frac{d}{dt}\boldsymbol{j} = \boldsymbol{m}_{\text{ext}} \tag{45}$$

where $\boldsymbol{j} \in \mathbb{R}^3$ is the total angular momentum of the Cosserat point and $\boldsymbol{m}_{\text{ext}} \in \mathbb{R}^3$ is the resultant external torque acting on the Cosserat point:

$$\begin{aligned} \boldsymbol{j} &= M\bar{\boldsymbol{x}} \times \dot{\bar{\boldsymbol{x}}} + E_0^{ij}\boldsymbol{d}_i \times \dot{\boldsymbol{d}}_j \\ \boldsymbol{m}_{\text{ext}} &= \bar{\boldsymbol{x}} \times \boldsymbol{f}_{\text{ext}} + \boldsymbol{d}_i \times \boldsymbol{f}^i_{\text{ext}} \end{aligned} \tag{46}$$

Note that both quantities are referred to the origin of the inertial frame of reference.

2.5 Balance of Energy

The balance law for energy can be obtained from the variational formulation (39) by substituting $\dot{\bar{\boldsymbol{x}}}$ for $\delta\bar{\boldsymbol{x}}$ and $\dot{\boldsymbol{d}}_i$ for $\delta\boldsymbol{d}_i$. Accordingly, we get

$$\dot{\bar{\boldsymbol{x}}} \cdot \left(M\ddot{\bar{\boldsymbol{x}}} - \boldsymbol{f}_{\text{ext}}\right) + \dot{\boldsymbol{d}}_i \cdot \left(E_0^{ij}\ddot{\boldsymbol{d}}_j + \frac{\partial U}{\partial \boldsymbol{d}_i} - \boldsymbol{f}^i_{\text{ext}}\right) = 0$$

where (33) has been employed. The last equation can be recast in the form

$$\frac{d}{dt}E = P_{\text{ext}} \tag{47}$$

where

$$P_{\text{ext}} = \boldsymbol{f}_{\text{ext}} \cdot \dot{\bar{\boldsymbol{x}}} + \boldsymbol{f}^i_{\text{ext}} \cdot \dot{\boldsymbol{d}}_i \tag{48}$$

denotes the power of the external forces acting on the body. Moreover, E is the total mechanical energy[1] given by

$$E = T + U \tag{49}$$

where U denotes the total strain energy defined in (13) and

$$T = \frac{1}{2}M\dot{\bar{\boldsymbol{x}}} \cdot \dot{\bar{\boldsymbol{x}}} + \frac{1}{2}E_0^{ij}\dot{\boldsymbol{d}}_i \cdot \dot{\boldsymbol{d}}_j \tag{50}$$

is the kinetic energy of the Cosserat point.

Remark 2.2 It is obvious from the balance law (45) that the total angular momentum is conserved (or a first integral of the motion) if the resultant external torque vanishes, that is, if $\boldsymbol{m}_{\text{ext}} = \boldsymbol{0}$. Then (43) yields

$$\begin{aligned} \boldsymbol{\xi} \cdot \left(M\bar{\boldsymbol{x}} \times \ddot{\bar{\boldsymbol{x}}} + E_0^{ij}\boldsymbol{d}_i \times \ddot{\boldsymbol{d}}_j\right) &= \boldsymbol{\xi} \cdot \left(\boldsymbol{d}_i \times \tfrac{\partial U}{\partial \boldsymbol{d}_i}\right) \\ \boldsymbol{\xi} \cdot \tfrac{d}{dt}\boldsymbol{j} &= \boldsymbol{\xi} \cdot \left(\overline{S}^{ij}\boldsymbol{d}_i \times \boldsymbol{d}_j\right) \\ &= 0 \end{aligned} \tag{51}$$

where use has been made of (33). Due to the arbitrariness of $\boldsymbol{\xi} \in \mathbb{R}^3$ the last equality implies that $\boldsymbol{j}$ is constant.

Remark 2.3 According to Noether's theorem conservation laws are intimately connected with invariance (or symmetry) properties of the system. In the present case conservation of angular momentum can be linked to the invariance of the potential energy under rotations.

[1] If the external loads or part of them can be derived from an associated potential energy function V_{ext} their contribution to the balance of energy can be shifted to the left-hand side of (47) by replacing U in (49) with $U + V_{\text{ext}}$.

In essence, the principle of material frame-indifference requires the stress response to be invariant under rigid motions. This requirement is satisfied by the fact that the total strain energy (13) is a function of the metric coefficients (29). That is,

$$U = \hat{U}(\boldsymbol{d}_i) = \tilde{U}(d_{ij}) \tag{52}$$

This implies invariance under rotations. To see this, let $\boldsymbol{d}_i^\sharp(t)$ define a motion that differs from $\boldsymbol{d}_i(t)$ by a rotation. Then, there is a rotation tensor $\boldsymbol{Q}(t) \in \mathrm{SO}(3)$ that belongs to the <u>S</u>pecial <u>O</u>rthogonal group in 3-space such that

$$\boldsymbol{d}_i^\sharp = \boldsymbol{Q}\boldsymbol{d}_i$$

It can be easily seen that the metric coefficients d_{ij} are invariant under rotations:

$$\begin{aligned} d_{ij}^\sharp &= \boldsymbol{d}_i^\sharp \cdot \boldsymbol{d}_j^\sharp \\ &= (\boldsymbol{Q}\boldsymbol{d}_i) \cdot \boldsymbol{Q}\boldsymbol{d}_j \\ &= \boldsymbol{d}_i \cdot \boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{d}_j \\ &= \boldsymbol{d}_i \cdot \boldsymbol{d}_j \\ &= d_{ij} \end{aligned} \tag{53}$$

With regard to (52) this implies rotational invariance of the total strain energy:

$$\hat{U}(\boldsymbol{Q}\boldsymbol{d}_i) = \hat{U}(\boldsymbol{d}_i) \tag{54}$$

Let $\boldsymbol{Q}_\varepsilon = \exp_{\mathrm{S0(3)}}(\varepsilon\widehat{\boldsymbol{\xi}}) \in \mathrm{SO}(3)$ for any $\varepsilon \in \mathbb{R}$ and skew-symmetric tensor $\widehat{\boldsymbol{\xi}} \in \mathrm{so}(3)$. In this connection, $\exp_{\mathrm{S0(3)}} : \mathrm{so}(3) \mapsto \mathrm{SO}(3)$ is the exponential map on the rotation group SO(3), given by the Rodrigues formula (see, for example, Marsden and Ratiu 1999)

$$\exp_{\mathrm{S0(3)}}(\varepsilon\widehat{\boldsymbol{\xi}}) = \boldsymbol{I} + \frac{\sin(\varepsilon\|\boldsymbol{\xi}\|)}{\|\boldsymbol{\xi}\|}\widehat{\boldsymbol{\xi}} + \frac{1}{2}\left[\frac{\sin(\varepsilon\|\boldsymbol{\xi}\|/2)}{\|\boldsymbol{\xi}/2\|}\right]^2 \widehat{\boldsymbol{\xi}}^2 \tag{55}$$

Here, $\widehat{\boldsymbol{\xi}} \in \mathrm{so}(3)$ is a skew-symmetric tensor with associated axial vector $\boldsymbol{\xi} \in \mathbb{R}^3$. That is, $\widehat{\boldsymbol{\xi}}\boldsymbol{a} = \boldsymbol{\xi} \times \boldsymbol{a}$ for any $\boldsymbol{a} \in \mathbb{R}^3$. It can be easily verified that $\boldsymbol{Q}_\varepsilon = \exp_{\mathrm{S0(3)}}(\varepsilon\widehat{\boldsymbol{\xi}})$ satisfies

$$\boldsymbol{Q}_{\varepsilon=0} = \boldsymbol{I} \quad \text{and} \quad \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \boldsymbol{Q}_\varepsilon = \widehat{\boldsymbol{\xi}}$$

Rotational invariance of the strain energy function (54) yields

$$\begin{aligned} 0 &= \tfrac{d}{d\varepsilon}\big|_{\varepsilon=0} \hat{U}\big(\boldsymbol{Q}_\varepsilon\boldsymbol{d}_i\big) \\ &= \tfrac{\partial\hat{U}}{\partial\boldsymbol{d}_i} \cdot \widehat{\boldsymbol{\xi}}\boldsymbol{d}_i \\ &= -\boldsymbol{\xi} \cdot \left(\tfrac{\partial\hat{U}}{\partial\boldsymbol{d}_i} \times \boldsymbol{d}_i\right) \end{aligned} \tag{56}$$

for any $\boldsymbol{\xi} \in \mathbb{R}^3$. Comparison of the last equation with (51)$_1$ shows that rotational invariance of the strain energy indeed yields the conservation law for angular momentum.

Remark 2.4 The rotational invariance of the total strain energy (52) is in agreement with Cauchy's representation theorem (see Truesdell and Noll 2004, Sect. 11, or Antman 2005, Chapter 8). Accordingly, if a scalar-valued function $\hat{f}(\boldsymbol{d}_i)$ is invariant under the proper orthogonal group then it depends only on the set of invariants $\mathbb{S}(\eta) \cup \mathbb{T}(\eta)$, where η denotes the ordered set of directors $\eta = \{\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3\}$ and

$$\begin{aligned} \mathbb{S}(\eta) &= \{\boldsymbol{d}_i \cdot \boldsymbol{d}_j, 1 \leq i \leq j \leq 3\} \\ \mathbb{T}(\eta) &= \{(\boldsymbol{d}_1 \times \boldsymbol{d}_2) \cdot \boldsymbol{d}_3\} \end{aligned}$$

The fact that the metric coefficients $d_{ij} \in \mathbb{S}(\eta)$ corroborates that the total strain energy $\hat{U}(\boldsymbol{d}_i) = \tilde{U}(d_{ij})$ is invariant under rotations.

2.6 The Link to Finite Elements

The initial value problem (40) fits into the standard framework for semi-discrete mechanical systems resulting from a space discretization of nonlinear elastodynamics. The finite element method is commonly used to perform the discretization in space of continuum bodies. This results in the semi-discrete equations of motion which assume the standard form

$$\boldsymbol{M}\ddot{\boldsymbol{q}} + \boldsymbol{F}_{\text{int}}(\boldsymbol{q}) = \boldsymbol{F}_{\text{ext}}(\boldsymbol{q}) \tag{57}$$

The system of nonlinear second-order ODEs (57) is subject to the initial conditions $\boldsymbol{q}(0) = \boldsymbol{q}_0$ and $\dot{\boldsymbol{q}}(0) = \boldsymbol{v}_0$, where $\boldsymbol{q}_0, \boldsymbol{v}_0 \in \mathbb{R}^n$ are given. For the free hyperelastic Cosserat point we have $n = 12$ DOFs. In particular, the configuration vector $\boldsymbol{q} : [0, T] \to \mathbb{R}^n$ of the Cosserat point is given by

$$\boldsymbol{q} = \begin{bmatrix} \boldsymbol{q}_1 \\ \vdots \\ \boldsymbol{q}_N \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{x}} \\ \boldsymbol{d}_1 \\ \boldsymbol{d}_2 \\ \boldsymbol{d}_3 \end{bmatrix} \tag{58}$$

where N denotes the number of 'nodal' configuration vectors $\boldsymbol{q}_A \in \mathbb{R}^3$ needed to describe the finite-dimensional mechanical system at hand. Obviously, for the Cosserat point we have $N = 4$. Taking into account the above partition of the configuration vector $\boldsymbol{q} \in \mathbb{R}^{3N}$, the equations of motion (57) can be recast in the equivalent form

$$\sum_{A,B=1}^{N} \delta\boldsymbol{q}_A \cdot \big(M^{AB}\ddot{\boldsymbol{q}}_A + \nabla_{\boldsymbol{q}_A} V(\boldsymbol{q}) - \boldsymbol{F}_{\text{ext}}^A(\boldsymbol{q})\big) = 0 \tag{59}$$

for arbitrary $\delta \boldsymbol{q}_A \in \mathbb{R}^3$, $A = 1, \ldots, N$. Comparing (59) with (39) yields the mass matrix $\boldsymbol{M} \in \mathbb{R}^{3N\times 3N}$ pertaining to the Cosserat point

$$\boldsymbol{M} = \begin{bmatrix} M\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & E_0^{11}\boldsymbol{I} & E_0^{12}\boldsymbol{I} & E_0^{13}\boldsymbol{I} \\ \boldsymbol{0} & E_0^{21}\boldsymbol{I} & E_0^{22}\boldsymbol{I} & E_0^{23}\boldsymbol{I} \\ \boldsymbol{0} & E_0^{31}\boldsymbol{I} & E_0^{32}\boldsymbol{I} & E_0^{33}\boldsymbol{I} \end{bmatrix} \tag{60}$$

where $\boldsymbol{I}$ denotes the 3×3 identity matrix. Note that the mass matrix is constant, symmetric and positive-definite. In this connection we remark that, as is obvious from (27), $E_0^{ij} = E_0^{ji}$. Moreover, $V : \mathbb{R}^n \to \mathbb{R}$ is a potential energy function which, in the case of the Cosserat point, originates from the strain energy (13). In addition to the internal force vector $\boldsymbol{F}_{\text{int}}(\boldsymbol{q}) = \nabla V(\boldsymbol{q})$, the external force vector $\boldsymbol{F}_{\text{ext}}(\boldsymbol{q}) \in \mathbb{R}^n$ might represent configuration dependent (follower) loads. For the Cosserat point we have

$$\boldsymbol{F}_{\text{ext}} = \begin{bmatrix} \boldsymbol{F}^1_{\text{ext}} \\ \vdots \\ \boldsymbol{F}^N_{\text{ext}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_{\text{ext}} \\ \boldsymbol{f}^1_{\text{ext}} \\ \boldsymbol{f}^2_{\text{ext}} \\ \boldsymbol{f}^3_{\text{ext}} \end{bmatrix} \tag{61}$$

2.6.1 The 4-node Tetrahedral Element

One of the most frequently used low-order elements is the 4-node tetrahedral element and its two dimensional counterpart, the 3-node triangle. The present formulation of the Cosserat point is equivalent to the 4-node tetrahedral element (Fig. 3). Specifically, the configuration vector (58) of the Cosserat point, $\boldsymbol{q} \in \mathbb{R}^{12}$, can be directly connected to the four nodal position vectors, $\boldsymbol{x}_A \in \mathbb{R}^3$, $A \in \{1, 2, 3, 4\}$, characterizing the deformed configuration of the tetrahedral element. In view of the kinematic

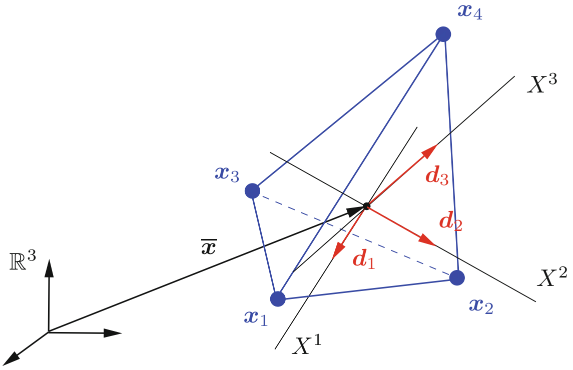


Fig. 3 The 4-node tetrahedral element and its connection with the Cosserat point

relationship (23), the nodal position vectors of the 4-node tetrahedral element can be expressed as

$$\boldsymbol{x}_A = \bar{\boldsymbol{x}} + X_A^i \boldsymbol{d}_i$$

where $\bar{\boldsymbol{x}} \in \mathbb{R}^3$ denotes the center of mass of the tetrahedral element and

$$X_A^i = \boldsymbol{D}^i \cdot (\boldsymbol{X}_A - \overline{\boldsymbol{X}})$$

are the material coordinates of the nodes in accordance with (22). We refer to Fig. 4 for an illustration of the planar case.

Introducing the nodal configuration vector of the 4-node tetrahedral element

$$\boldsymbol{q}^e = \left[\boldsymbol{x}_1^T \; \boldsymbol{x}_2^T \; \boldsymbol{x}_3^T \; \boldsymbol{x}_4^T\right]^T$$

we may write

$$\boldsymbol{q}^e = \boldsymbol{T}\boldsymbol{q}$$

where the configuration vector of the Cosserat point, $\boldsymbol{q} \in \mathbb{R}^{12}$ is given by (58), and $\boldsymbol{T}$ is a constant 12 × 12 transformation matrix of the form

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{I} & X_1^1\boldsymbol{I} & X_1^2\boldsymbol{I} & X_1^3\boldsymbol{I} \\ \boldsymbol{I} & X_2^1\boldsymbol{I} & X_2^2\boldsymbol{I} & X_2^3\boldsymbol{I} \\ \boldsymbol{I} & X_3^1\boldsymbol{I} & X_3^2\boldsymbol{I} & X_3^3\boldsymbol{I} \\ \boldsymbol{I} & X_4^1\boldsymbol{I} & X_4^2\boldsymbol{I} & X_4^3\boldsymbol{I} \end{bmatrix}$$

For regular geometries of the tetrahedral element, matrix $\boldsymbol{T}$ is non-singular. Substituting the relationships

$$\boldsymbol{q} = \boldsymbol{T}^{-1}\boldsymbol{q}^e\,, \quad \delta\boldsymbol{q} = \boldsymbol{T}^{-1}\delta\boldsymbol{q}^e\,, \quad \ddot{\boldsymbol{q}} = \boldsymbol{T}^{-1}\ddot{\boldsymbol{q}}^e$$

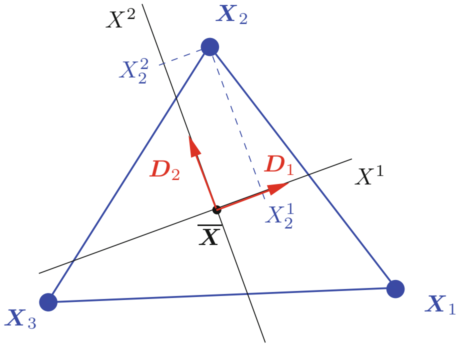


Fig. 4 The 3-node triangular element in the reference configuration. The position of node 2 relative to the center of mass, $\overline{\boldsymbol{X}} \in \mathbb{R}^2$, is characterized by the coordinates X_2^1 and X_2^2

into (59), the equations of motion can be written in terms of the nodal quantities pertaining to the tetrahedral element. For example, the corresponding mass matrix is given by

$$\boldsymbol{M}^e = \boldsymbol{T}^{-T}\boldsymbol{M}\boldsymbol{T}^{-1}$$

where the mass matrix $\boldsymbol{M}$ of the Cosserat point is given by (60).

2.7 The EM Method Due to Simo and Tarnow

We start our treatment of EM schemes by applying the method developed by Simo and Tarnow (1992) in the context of nonlinear elastodynamics to the model problem of an elastic Cosserat point. In essence the discretization in time of the ODEs (40) consists of a modification of the mid-point rule. Correspondingly, the resulting EM integrator is an implicit second-order scheme. For the present purposes it is convenient to confine our attention to the balances of director momentum $(40)_2$. Accordingly, we shall consider the following initial value problem written in first-order form: Find $\boldsymbol{d}_i, \boldsymbol{v}_i : [0, T] \rightarrow \mathbb{R}^3$, $(i = 1, 2, 3)$, such that

$$\begin{aligned} \dot{\boldsymbol{d}}_i &= \boldsymbol{v}_i \\ E_0^{ij}\dot{\boldsymbol{v}}_j &= \boldsymbol{f}^i_{\text{ext}} - \overline{S}^{ij}\boldsymbol{d}_j \end{aligned} \tag{62}$$

subject to the initial conditions $\boldsymbol{d}_i(0) = (\boldsymbol{d}_i)_0$, and $\boldsymbol{v}_i(0) = (\boldsymbol{v}_i)_0$, where $(\boldsymbol{d}_i)_0, (\boldsymbol{v}_i)_0 \in \mathbb{R}^3$ are given quantities. Note that expression (33) for the internal director forces, $\boldsymbol{f}^i_{\text{int}} = \overline{S}^{ij}\boldsymbol{d}_j$, has been used in $(62)_2$. Consider a representative time interval $[t_n, t_{n+1}]$ with time step $\Delta t = t_{n+1} - t_n$, and given state-space coordinates $\boldsymbol{d}_{i_n} \in \mathbb{R}^3$ and $\boldsymbol{v}_{i_n} \in \mathbb{R}^3$ at t_n. The resulting algebraic problem to be solved is given as follows: Find $(\boldsymbol{d}_{i_{n+1}}, \boldsymbol{v}_{i_{n+1}}) \in \mathbb{R}^3 \times \mathbb{R}^3$, $(i = 1, 2, 3)$, as the solution of the algebraic system of equations

$$\begin{aligned} \boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n} &= \Delta t \boldsymbol{v}_{i_{n+\frac{1}{2}}} \\ E_0^{ij}\big(\boldsymbol{v}_{j_{n+1}} - \boldsymbol{v}_{j_n}\big) &= \Delta t\big(\,(\boldsymbol{f}^i_{\text{ext}})\big|_{n+\frac{1}{2}} - \overline{S}^{ij}_A\boldsymbol{d}_{j_{n+\frac{1}{2}}}\big) \end{aligned} \tag{63}$$

Here and in the sequel $(\bullet)_{n+\frac{1}{2}}$ denotes the mean value of the quantity $(\bullet)$ in the time interval $[t_n, t_{n+1}]$. That is,

$$(\bullet)_{n+\frac{1}{2}} = \frac{1}{2}\big((\bullet)_n + (\bullet)_{n+1}\big) \tag{64}$$

Moreover, $(\boldsymbol{f}^i_{\text{ext}})\big|_{n+\frac{1}{2}}$ denotes the approximation of the external director forces in the time interval $[t_n, t_{n+1}]$, the specification of which is left open at the present stage. The distinguishing feature of the scheme (63) is the presence of an algorithmic stress formula for the calculation of $\overline{S}^{ij}_A$. In particular, for the specific case of *St. Venant-*

Kirchhoff material, in Simo and Tarnow (1992) the following closed-form expression for $\overline{S}_A^{ij}$ is proposed:

$$\overline{S}_A^{ij} = \mathsf{C}^{ijkl}\gamma_{kl_{n+\frac{1}{2}}} \tag{65}$$

Here, $\mathsf{C}^{ijkl} = 4\frac{\partial^2 U}{\partial d_{ij}\partial d_{kl}}$ are the components of the fourth-order elasticity tensor and γ_{ij} denote the components of the Green–Lagrangean strain tensor given by

$$\gamma_{ij} = \frac{1}{2}\big(d_{ij} - \delta_{ij}\big) \tag{66}$$

In view of (64) and (66), the algorithmic stress formula (65) relies on the mean value of the metric coefficients

$$\begin{aligned} d_{ij_{n+\frac{1}{2}}} &= \tfrac{1}{2}\left(d_{ij_n} + d_{ij_{n+1}}\right) \\ &= \tfrac{1}{2}\left(\boldsymbol{d}_{i_n}\cdot\boldsymbol{d}_{j_n} + \boldsymbol{d}_{i_{n+1}}\cdot\boldsymbol{d}_{j_{n+1}}\right) \end{aligned}$$

Note that this is in contrast to the mid-point rule, in which

$$\begin{aligned} d_{ij}^{\text{MP}} &= \boldsymbol{d}_{i_{n+\frac{1}{2}}}\cdot\boldsymbol{d}_{j_{n+\frac{1}{2}}} \\ &= \tfrac{1}{2}d_{ij_{n+\frac{1}{2}}} + \tfrac{1}{4}\left(\boldsymbol{d}_{i_n}\cdot\boldsymbol{d}_{j_{n+1}} + \boldsymbol{d}_{i_{n+1}}\cdot\boldsymbol{d}_{j_n}\right) \\ &= d_{ij_{n+\frac{1}{2}}} - \tfrac{\Delta t^2}{4}\boldsymbol{v}_{i_{n+\frac{1}{2}}}\cdot\boldsymbol{v}_{j_{n+\frac{1}{2}}} \end{aligned} \tag{67}$$

would have to be used. In the last equation use has been made of $(63)_1$. We next show that the scheme is capable of conserving both angular momentum and energy.

2.7.1 Algorithmic Conservation of Angular Momentum

With regard to $(46)_1$ the angular momentum $\bar{\boldsymbol{j}}$ of the Cosserat point relative to its center of mass can be written in the form.

$$\bar{\boldsymbol{j}}(\boldsymbol{d}_i, \boldsymbol{v}_i) = E_0^{ij}\boldsymbol{d}_i \times \boldsymbol{v}_j \tag{68}$$

Obviously, the angular momentum is a quadratic function of the state-space coordinates $(\boldsymbol{d}_i, \boldsymbol{v}_i)$. To calculate the incremental change in the angular momentum we take into account the following remark.

Remark 2.5 When the map $f : \mathbb{R}^k \to \mathbb{R}$ is at most quadratic then the relationship

$$Df(\boldsymbol{y}_{n+\frac{1}{2}})\cdot(\boldsymbol{y}_{n+1} - \boldsymbol{y}_n) = f(\boldsymbol{y}_{n+1}) - f(\boldsymbol{y}_n) \tag{69}$$

holds.

Accordingly, setting $\boldsymbol{y} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_3, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_3)$, we get

$$\begin{aligned}\bar{\boldsymbol{j}}_{n+1} - \bar{\boldsymbol{j}}_n &= \bar{\boldsymbol{j}}(\boldsymbol{y}_{n+1}) - \bar{\boldsymbol{j}}(\boldsymbol{y}_n)\\ &= \tfrac{\partial \bar{\boldsymbol{j}}}{\partial \boldsymbol{d}_i}(\boldsymbol{y}_{n+\frac{1}{2}})(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}) + \tfrac{\partial \bar{\boldsymbol{j}}}{\partial \boldsymbol{v}_i}(\boldsymbol{y}_{n+\frac{1}{2}})(\boldsymbol{v}_{i_{n+1}} - \boldsymbol{v}_{i_n})\\ &= -E_0^{ij}\widehat{\boldsymbol{v}}_{j_{n+\frac{1}{2}}}(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}) + E_0^{ji}\widehat{\boldsymbol{d}}_{j_{n+\frac{1}{2}}}(\boldsymbol{v}_{i_{n+1}} - \boldsymbol{v}_{i_n})\end{aligned}$$

Substituting form (63)$_1$ and (63)$_2$ into the last equation and taking into account the symmetry property $E_0^{ij} = E_0^{ji}$ yields

$$\begin{aligned}\bar{\boldsymbol{j}}_{n+1} - \bar{\boldsymbol{j}}_n &= \widehat{\boldsymbol{d}}_{j_{n+\frac{1}{2}}} \Delta t\big((\boldsymbol{f}^j_{\text{ext}})\Big|_{n+\frac{1}{2}} - \overline{S}_A^{ji}\boldsymbol{d}_{i_{n+\frac{1}{2}}}\big)\\ &= \Delta t \boldsymbol{d}_{j_{n+\frac{1}{2}}} \times (\boldsymbol{f}^j_{\text{ext}})\Big|_{n+\frac{1}{2}}\\ &= \Delta t\, \overline{\boldsymbol{m}}_{\text{ext}}|_{n+\frac{1}{2}}\end{aligned} \tag{70}$$

where the symmetry of $\overline{S}_A^{ji}$ has been accounted for. In the last equation $\overline{\boldsymbol{m}}_{\text{ext}}|_{n+\frac{1}{2}}$ denotes the discrete version of the resultant external torque relative to the center of mass defined by

$$\overline{\boldsymbol{m}}_{\text{ext}} = \boldsymbol{d}_i \times \boldsymbol{f}^i_{\text{ext}} \tag{71}$$

Note that this definition is in line with (46)$_2$. It is obvious from (70) that the present scheme conserves the angular momentum provided that the external torque vanishes.

2.7.2 Algorithmic Conservation of Energy

Combining (63)$_1$ and (63)$_2$ using the dot product leads to

$$\big(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\big) \cdot \big((\boldsymbol{f}^i_{\text{ext}})\big|_{n+\frac{1}{2}} - \overline{S}_A^{ij}\boldsymbol{d}_{j_{n+\frac{1}{2}}}\big) = \boldsymbol{v}_{i_{n+\frac{1}{2}}} \cdot E_0^{ij}\big(\boldsymbol{v}_{j_{n+1}} - \boldsymbol{v}_{j_n}\big) \tag{72}$$

Concerning the right-hand side of the last equation we get

$$\begin{aligned}\boldsymbol{v}_{i_{n+\frac{1}{2}}} \cdot E_0^{ij}\big(\boldsymbol{v}_{j_{n+1}} - \boldsymbol{v}_{j_n}\big) &= \tfrac{1}{2}E_0^{ij}\big(\boldsymbol{v}_{i_{n+1}} \cdot \boldsymbol{v}_{j_{n+1}} - \boldsymbol{v}_{i_n} \cdot \boldsymbol{v}_{j_n}\big)\\ &= \overline{T}_{n+1} - \overline{T}_n\end{aligned}$$

where the symmetry of E_0^{ij} has been taken into account. Moreover, $\overline{T}$ denotes the relative kinetic energy given by

$$\overline{T} = \frac{1}{2}E_0^{ij}\boldsymbol{v}_i \cdot \boldsymbol{v}_j \tag{73}$$

Note that the above expression for the relative kinetic energy is in line with definition (50) of the total kinetic energy of the Cosserat point. Furthermore,

$$\begin{aligned}\big(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\big) \cdot \overline{S}_A^{ij} \boldsymbol{d}_{j_{n+\frac{1}{2}}} &= \tfrac{1}{2}\overline{S}_A^{ij}\big(\boldsymbol{d}_{i_{n+1}} \cdot \boldsymbol{d}_{j_{n+1}} - \boldsymbol{d}_{i_n} \cdot \boldsymbol{d}_{j_n}\big)\\ &= \tfrac{1}{2}\overline{S}_A^{ij}\big(d_{ij_{n+1}} - d_{ij_n}\big)\\ &= \overline{S}_A^{ij}\big(\gamma_{ij_{n+1}} - \gamma_{ij_n}\big)\end{aligned}$$

where use has been made of the symmetry of $\overline{S}_A^{ij}$ along with the definition of the metric coefficients and the Green–Lagrangean strains (66). Employing the algorithmic stress formula (65), the last equation gives

$$\begin{aligned}\overline{S}_A^{ij}\big(\gamma_{ij_{n+1}} - \gamma_{ij_n}\big) &= \mathsf{C}^{ijkl}\gamma_{kl_{n+\frac{1}{2}}}\big(\gamma_{ij_{n+1}} - \gamma_{ij_n}\big)\\ &= \tfrac{1}{2}\big(\gamma_{ij_{n+1}} - \gamma_{ij_n}\big)\mathsf{C}^{ijkl}\big(\gamma_{kl_{n+1}} + \gamma_{kl_n}\big)\\ &= \tfrac{1}{2}\big(\gamma_{ij_{n+1}}\mathsf{C}^{ijkl}\gamma_{kl_{n+1}} - \gamma_{ij_n}\mathsf{C}^{ijkl}\gamma_{kl_n}\big)\end{aligned}$$

where the major symmetry $\mathsf{C}^{ijkl} = \mathsf{C}^{klij}$ of the elasticity tensor has been taken into account. The total strain energy of the St. Venant-Kirchhoff Cosserat point is given by

$$U^{\mathrm{St.V-K}} = \frac{1}{2}\gamma_{ij}\mathsf{C}^{ijkl}\gamma_{kl} \tag{74}$$

Altogether, Eq. (72) can be recast in the form

$$\begin{aligned}\Delta t\ (\boldsymbol{f}^i_{\mathrm{ext}})\big|_{n+\frac{1}{2}}\ \boldsymbol{v}_{i_{n+\frac{1}{2}}} &= \overline{T}_{n+1} - \overline{T}_n + U^{\mathrm{St.V-K}}_{n+1} - U^{\mathrm{St.V-K}}_n\\ &= \overline{E}_{n+1} - \overline{E}_n\end{aligned} \tag{75}$$

where on the left-hand side of the last equation use has been made of $(63)_1$. Moreover, on the right-hand side the total energy $\overline{E}$ has been introduced, analogous to (49). The last equation corroborates algorithmic conservation of energy in the absence of external loading.

Remark 2.6 The above investigation shows that in order to achieve algorithmic conservation of energy the stress formula has to satisfy the condition

$$U_{n+1} - U_n = \frac{1}{2}\overline{S}_A^{ij}\big(d_{ij_{n+1}} - d_{ij_n}\big) \tag{76}$$

This condition can be viewed as discrete counterpart of

$$\frac{d}{dt}\tilde{U}(d_{ij}) = \frac{\partial\tilde{U}}{\partial d_{ij}}\dot{d}_{ij} = \frac{1}{2}\overline{S}^{ij}\dot{d}_{ij}$$

In the last equation use has been made of (31).

Remark 2.7 The nonlinear system of equations emanating from the EM scheme (63) is typically solved iteratively by applying Newton's method. As outlined in Appendix A.3, the corresponding iteration matrix is nonsymmetric. This is in contrast to standard schemes such as the mid-point rule which yield a symmetric iteration matrix. This is the price one has to pay for the improved numerical stability of the EM integrator.

2.8 The Discrete Derivative

The closed-form expression for the algorithmic stress formula (65) proposed by Simo and Tarnow (1992) is restricted to St. Venant-Kirchhoff material. A generalized procedure for the design of second-order EM integrators relies on the notation of a discrete derivative introduced in Gonzalez (1996). This approach makes possible the design of appropriate algorithmic stress formulas for general hyperelastic constitutive laws. Moreover, symmetries of the mechanical system can be taken into account by the introduction of specific invariants. If the invariants and the corresponding momentum maps are *at most quadratic*, the resulting time-stepping scheme is capable of conserving the respective momentum map. As has been shown above the momentum map of primary interest in the present work is the total angular momentum. In this connection the metric coefficients play the role of quadratic invariants (see Remark 2.4). In analogy to (35), see Remark 2.1, the discrete version of the derivative $\partial \hat{U}/\partial \boldsymbol{d}_i$ is chosen to be

$$\overline{\nabla}_{d_i} U(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) = \mathsf{D}\tilde{U}(d_{jk_n}, d_{jk_{n+1}}) \frac{\partial d_{jk}}{\partial \boldsymbol{d}_i}(\boldsymbol{d}_{l_{n+\frac{1}{2}}}) \tag{77}$$

where

$$\mathsf{D}\tilde{U}(d_{ik_n}, d_{ik_{n+1}}) = \mathsf{S}^{ik} + \frac{U_{n+1} - U_n - \mathsf{S}^{jl}\Delta d_{jl}}{\Delta d_{mn}\Delta d_{mn}}\Delta d_{ik} \tag{78}$$

with

$$\mathsf{S}^{ik} = D\tilde{U}(d_{ik_{n+\frac{1}{2}}}) \qquad \text{and} \qquad \Delta d_{ik} = d_{ik_{n+1}} - d_{ik_n}$$

Similar to (35), the discrete gradient (77) can be written as

$$\overline{\nabla}_{d_i} U(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) = 2\mathsf{D}\tilde{U}(d_{ik_n}, d_{ik_{n+1}})\boldsymbol{d}_{k_{n+\frac{1}{2}}} \tag{79}$$

Using the discrete gradient (77), the EM scheme (63) can be recast in the form

$$\begin{aligned} \boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n} &= \Delta t \boldsymbol{v}_{i_{n+\frac{1}{2}}} \\ E_0^{ij}\big(\boldsymbol{v}_{j_{n+1}} - \boldsymbol{v}_{j_n}\big) &= \Delta t\big((\boldsymbol{f}_{\text{ext}}^i)\big|_{n+\frac{1}{2}} - \overline{\nabla}_{d_i} U(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}})\big) \end{aligned} \tag{80}$$

We refer to this scheme as the *EM integrator for hyperelastic Cosserat points*. In the present context the discrete gradient (79) gives rise to the algorithmic stress formula

$$\overline{S}^{ij}_A = 2\mathsf{D}\tilde{U}(d_{ij_n}, d_{ij_{n+1}}) \tag{81}$$

2.8.1 Directionality Property of the Discrete Derivative

In analogy to the following continuous relationship

$$\frac{d}{dt}U = \frac{\partial \hat{U}}{\partial \boldsymbol{d}_k} \cdot \dot{\boldsymbol{d}}_k$$

the discrete derivative satisfies by design the so-called directionality property (Gonzalez 1996)

$$U_{n+1} - U_n = \overline{\nabla}_{d_i} U \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_{n+1}}\right) \tag{82}$$

To see this, substitute from (77) into the last equation to get

$$\begin{aligned} U_{n+1} - U_n &= \mathsf{D}\tilde{U}(d_{jk_n}, d_{jk_{n+1}}) \tfrac{\partial d_{jk}}{\partial \boldsymbol{d}_i}(\boldsymbol{d}_{l_{n+\frac{1}{2}}}) \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_{n+1}}\right) \\ &= \mathsf{D}\tilde{U}(d_{jk_n}, d_{jk_{n+1}})\left(d_{jk_{n+1}} - d_{jk_n}\right) \end{aligned} \tag{83}$$

Here Remark 2.5 has been applied since the metric coefficients d_{jk} are merely quadratic functions of $\boldsymbol{d}_i$. It can be easily verified that formula (78) satisfies the last equation by design.

2.8.2 Algorithmic Conservation Properties

Due to the directionality property $(83)_2$ the algorithmic stress formula (81) automatically fulfills the condition (76) for the conservation of energy. Moreover, the above proof of algorithmic conservation of angular momentum remains unaltered.

Remark 2.8 Condition (76) for algorithmic energy conservation can be used as algebraic constraint in an optimization problem to devise suitable stress formulas for second-order EM schemes, see Groß et al. (2005, Sect. 6.8) and Romero (2012). In particular, in these works, formula (78) is derived by applying the optimization approach. Moreover, the optimization approach is employed in the Galerkin-based discretization method in Groß et al. (2005, Sect. 6) to construct higher-order EM schemes for nonlinear elastodynamics.

Remark 2.9 While the notion of a discrete derivative makes possible the design of EM schemes for general hyperelastic constitutive laws, stress formula (81) boils down to (65) in the case of the St. Venant-Kirchhoff model. This can be shown by inserting the total strain energy (74) pertaining to the St. Venant-Kirchhoff model into the discrete derivative (78).

Remark 2.10 The discrete derivative of a quadratic function coincides with the standard derivative evaluated at $(\bullet)_{n+\frac{1}{2}}$. In particular, since the strain energy of the St. Venant-Kirchhoff model is merely a quadratic function of the metric coefficients or Green–Lagrangean strains, respectively, the discrete derivative (81) coincides with the standard derivative

$$\begin{aligned} \overline{S}_A^{ij} &= 2D\tilde{U}^{\mathrm{St.V-K}}\big(d_{ij_{n+\frac{1}{2}}}\big) \\ &= D\tilde{U}^{\mathrm{St.V-K}}\big(\gamma_{ij_{n+\frac{1}{2}}}\big) \\ &= \mathsf{C}^{ijkl}\gamma_{kl_{n+\frac{1}{2}}} \end{aligned} \tag{84}$$

where relation (66) between the metric coefficients d_{ij} and the Green–Lagrangean strains γ_{ij} has been taken into account along with the strain energy function (74). This result is in agreement with stress formula (65) due to Simo and Tarnow (1992), and therefore complements Remark 2.9.

Remark 2.11 In the *linearized theory* the strains are merely linear functions of the displacements. In the present context the linearized strains are given by

$$\gamma_{ij}^{\mathrm{lin}} = \frac{1}{2}\left(\boldsymbol{u}_i \cdot \boldsymbol{D}_j + \boldsymbol{D}_i \cdot \boldsymbol{u}_j\right)$$

Note that the director displacements $\boldsymbol{u}_i \in \mathbb{R}^3$ have been introduced such that the relationship $\boldsymbol{d}_i = \boldsymbol{D}_i + \boldsymbol{u}_i$ holds. The strain energy (74) thus becomes a quadratic function of the displacements and can be written as

$$U_{\mathrm{lin}} = \frac{1}{2}\boldsymbol{u}_i \cdot \boldsymbol{K}^{ij}\boldsymbol{u}_j$$

where $\boldsymbol{K}^{ij} \in \mathbb{R}^{3\times 3}$ constitutes a symmetric stiffness matrix. Consequently, due to the properties of the discrete derivative (cf. Remark 2.10),

$$\begin{aligned} \overline{\nabla}_{d_i} U_{\mathrm{lin}}(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) &= \nabla U_{\mathrm{lin}}(\boldsymbol{d}_{j_{n+\frac{1}{2}}}) \\ &= \tfrac{1}{2}\left(\nabla U_{\mathrm{lin}}(\boldsymbol{d}_{j_n}) + \nabla U_{\mathrm{lin}}(\boldsymbol{d}_{j_{n+1}})\right) \\ &= \tfrac{1}{2}\boldsymbol{K}^{ij}\left(\boldsymbol{u}_{j_n} + \boldsymbol{u}_{j_{n+1}})\right) \end{aligned}$$

Accordingly, for linear problems the EM integrator (80) coincides with the trapezoidal rule (or average acceleration method) which is a member of the Newmark family, see Hughes (2000). The average acceleration method is known to be energy preserving, unconditionally stable, and one of the most widely used methods for structural dynamics applications.

3 Rigid Body Dynamics

3.1 *From the Cosserat Point to the Rigid Body*

We next perform the transition from the theory of a Cosserat point to rigid body dynamics (Fig. 5). To this end we consider the imposition of geometric constraints on the Cosserat point. In particular, the constraints can be included in a straightforward way by replacing the strain energy (13) with an augmented potential function

$$V_\lambda(\boldsymbol{d}_i) = \hat{U}(\boldsymbol{d}_i) + \sum_{l=1}^{R} \lambda^l \hat{g}_l(\boldsymbol{d}_i) \tag{85}$$

Here, $\lambda^l : [0, T] \to \mathbb{R}$ are Lagrange multipliers for the enforcement of the (holonomic) constraints $g_l = 0$. Similar to the strain energy (52), frame-indifferent constraint functions are given by

$$g_l = \hat{g}_l(\boldsymbol{d}_i) = \tilde{g}_l(d_{ij}) \tag{86}$$

For example, the constraint $g_1 = \boldsymbol{d}_1 \cdot \boldsymbol{d}_1 - 1 = 0$ eliminates extension in the direction of $\boldsymbol{d}_1$. For a rigid body we have to impose $R = 6$ independent constraints. To this end we choose (85) to be of the form

$$\begin{aligned} V_\lambda^{\mathrm{RB}} = \sum_{l=1}^{6} \lambda^l \tilde{g}_l(d_{ij}) &= \boldsymbol{\Lambda} : \tfrac{1}{2}(\boldsymbol{C} - \boldsymbol{I}) \\ &= \boldsymbol{\Lambda} : \tfrac{1}{2}\big((d_{ij} - \delta_{ij})\boldsymbol{D}^i \otimes \boldsymbol{D}^j\big) \\ &= \boldsymbol{D}^i \cdot \boldsymbol{\Lambda}\boldsymbol{D}^j \tfrac{1}{2}(d_{ij} - \delta_{ij}) \\ &= \Lambda^{ij} \tfrac{1}{2}(d_{ij} - \delta_{ij}) \end{aligned} \tag{87}$$

Here, the Lagrange multipliers are contained in the symmetric tensor $\boldsymbol{\Lambda}$ according to the following assignment

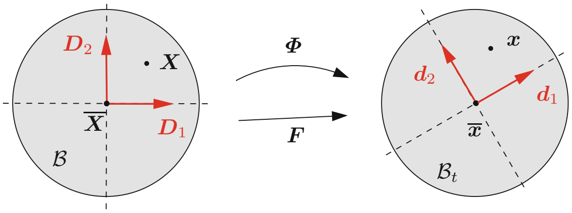


Fig. 5 Planar illustration of the transition from the elastic Cosserat point to the rigid body: The director frame $\{\boldsymbol{d}_i\}$ is forced to stay orthonormal for all time. Correspondingly, $\boldsymbol{F} \in \mathrm{SO}(3)$, see Remark 3.1

$$\begin{bmatrix} \lambda^1 \\ \lambda^2 \\ \lambda^3 \\ \lambda^4 \\ \lambda^5 \\ \lambda^6 \end{bmatrix} = \begin{bmatrix} \Lambda^{11} \\ \Lambda^{22} \\ \Lambda^{33} \\ \Lambda^{12} \\ \Lambda^{13} \\ \Lambda^{23} \end{bmatrix} \tag{88}$$

giving rise to the following six independent constraints of rigidity:

$$[\tilde{g}_l(d_{ij})] = \begin{bmatrix} \frac{1}{2}(d_{11} - 1) \\ \frac{1}{2}(d_{22} - 1) \\ \frac{1}{2}(d_{33} - 1) \\ d_{12} \\ d_{13} \\ d_{23} \end{bmatrix} = \mathbf{0} \tag{89}$$

As before (see Sect. 2.2), we assume that the director triad $\{\boldsymbol{D}_i\}$ in the reference configuration is orthonormal, that is, $\boldsymbol{D}_i \cdot \boldsymbol{D}_j = \delta_{ij}$. Accordingly, the constraints $\tilde{g}_l(\delta_{ij}) = 0$ ($l = 1, \ldots, 6$) are identically fulfilled in the reference configuration. Imposing these constraints forces the director triad $\{\boldsymbol{d}_i(t)\}$ to stay orthonormal for all time. The configuration space corresponding to the motion of the rigid body about its center of mass is given by

$$\mathsf{Q} = \{\boldsymbol{d}_i \in \mathbb{R}^3 \mid \tilde{g}_l(d_{ij}) = 0\,,\ \ 1 \leq l \leq 6\,, (\boldsymbol{d}_1 \times \boldsymbol{d}_2) \cdot \boldsymbol{d}_3 = 1\} \tag{90}$$

Accordingly, the nine director components are subject to six independent constraints of rigidity. This is in agreement with the fact that the rotational motion of a rigid body has three degrees of freedom. The director velocities $\boldsymbol{v}_i$ have to belong to the tangent space to Q at $\boldsymbol{d}_i \in \mathsf{Q}$ given by

$$T_{d_i}\mathsf{Q} = \{\boldsymbol{v}_i \in \mathbb{R}^3 \mid \boldsymbol{v}_i = \boldsymbol{\omega} \times \boldsymbol{d}_i\,, \boldsymbol{\omega} \in \mathbb{R}^3\} \tag{91}$$

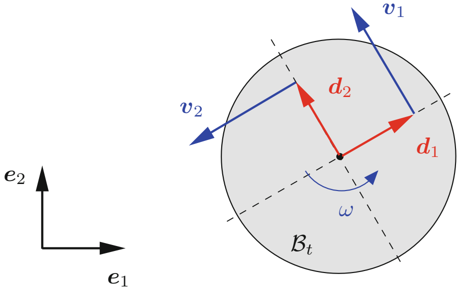
where $\boldsymbol{\omega}$ is the angular velocity (Fig. 6). It can be easily verified that the director velocities $\boldsymbol{v}_i \in T_{d_i}\mathsf{Q}$ satisfy the constraints on the velocity level given by

$$\frac{d}{dt}\left(\frac{1}{2}\big(d_{ij} - \delta_{ij}\big)\right) = \frac{1}{2}\big(\boldsymbol{d}_i \cdot \boldsymbol{v}_j + \boldsymbol{v}_i \cdot \boldsymbol{d}_j\big) = 0 \tag{92}$$

The equations governing the rotational motion of the free rigid body can be easily deduced from the Cosserat point by replacing the internal director forces

$$\boldsymbol{f}^i_{\text{int}} = \frac{\partial U}{\partial \boldsymbol{d}_i} = 2\frac{\partial U}{\partial d_{ik}}\boldsymbol{d}_k = \overline{S}^{ik}\boldsymbol{d}_k \tag{93}$$

Fig. 6 Planar illustration of the rotation of a rigid body with angular velocity $\boldsymbol{\omega} = \omega \boldsymbol{e}_3$ and director velocities $\boldsymbol{v}_i \in T_{d_i}\mathsf{Q}$



with the constraint director forces

$$\boldsymbol{f}_{\mathrm{c}}^{i} = \frac{\partial V_{\lambda}^{\mathrm{RB}}}{\partial \boldsymbol{d}_i} = \sum_{l=1}^{6} \lambda^l \frac{\partial g_l}{\partial \boldsymbol{d}_i} = 2\frac{\partial V_{\lambda}^{\mathrm{RB}}}{\partial d_{ik}} \boldsymbol{d}_k = \Lambda^{ik} \boldsymbol{d}_k \tag{94}$$

Here, Remark 2.1 along with (87) have been taken into account. Now the initial value problem governing the rotational motion of the rigid body can be directly deduced from the corresponding problem pertaining to the Cosserat point (see Sect. 2.7): Find $(\boldsymbol{d}_i, \boldsymbol{v}_i) \in \mathbb{R}^3 \times \mathbb{R}^3$ $(i = 1, 2, 3)$, and $\lambda_l \in \mathbb{R}$ $(l = 1, \ldots, 6)$, such that

$$\begin{aligned} \dot{\boldsymbol{d}}_i &= \boldsymbol{v}_i \\ E_0^{ij} \dot{\boldsymbol{v}}_j &= \boldsymbol{f}_{\mathrm{ext}}^{i} - \Lambda^{ij} \boldsymbol{d}_j \\ \hat{g}_l(\boldsymbol{d}_i) &= 0 \end{aligned} \tag{95}$$

subject to the initial conditions $\boldsymbol{d}_i(0) = (\boldsymbol{d}_i)_0$, and $\boldsymbol{v}_i(0) = (\boldsymbol{v}_i)_0$, where $(\boldsymbol{d}_i)_0 \in \mathsf{Q}$, and $(\boldsymbol{v}_i)_0 \in T_{d_i}\mathsf{Q}$ are given quantities. While the motion of the hyperelastic Cosserat point is governed by ODEs, the present rigid body formulation relies on differential-algebraic equations (DAEs). As is common with constrained mechanical systems the DAEs (95) have (differential) index three. For more background on DAEs we refer to Ascher and Petzold (1998) and Kunkel and Mehrmann (2006). Note that the constraints $(95)_3$ provide six algebraic equations for the determination of the six independent Lagrange multipliers (88). In contrast to that, the six stress resultants $\overline{S}^{ij}$ of the hyperelastic Cosserat point are depending on the Green–Lagrangean strains (or metric coefficients) via the constitutive law.

Remark 3.1 Imposition of the six independent constraints of rigidity (89) is equivalent to the enforcement of zero Green–Lagrangean strains, that is $\gamma_{ij} = 0$, or

$$\boldsymbol{G} = \frac{1}{2}(\boldsymbol{C} - \boldsymbol{I}) = \gamma_{ij} \boldsymbol{D}^i \otimes \boldsymbol{D}^j = \boldsymbol{0}$$

The last equation implies

$$\boldsymbol{C} = \boldsymbol{F}^T \boldsymbol{F} = \boldsymbol{I}$$

Taking into account the original requirement $\det(\boldsymbol{F}) > 0$ for the Cosserat point, the last equation yields

$$\boldsymbol{F}^T\boldsymbol{F} = \boldsymbol{I} \qquad \text{and} \qquad \det(\boldsymbol{F}) = 1$$

Consequently, in the case of the rigid body, the deformation gradient coincides with a rotation tensor $\boldsymbol{F} \in \mathrm{SO}(3)$.

3.2 Balance of Angular Momentum

For the free rigid body, balance of angular momentum can be shown along the lines of the previous treatment of the Cosserat point. In particular, the balance law (45) together with the definition of the angular momentum and the resultant external torque in (46) remain unaltered. Accordingly, scalar multiplying $(95)_2$ by $\boldsymbol{\xi} \times \boldsymbol{d}_i$ yields

$$(\boldsymbol{\xi} \times \boldsymbol{d}_i) \cdot E_0^{ij}\dot{\boldsymbol{v}}_j = (\boldsymbol{\xi} \times \boldsymbol{d}_i) \cdot (\boldsymbol{f}^i_{\text{ext}} - \Lambda^{ij}\boldsymbol{d}_j)$$

or

$$\boldsymbol{\xi} \cdot \Big(E_0^{ij}\boldsymbol{d}_i \times \dot{\boldsymbol{v}}_j - \boldsymbol{d}_i \times \boldsymbol{f}^i_{\text{ext}} + \Lambda^{ij}\boldsymbol{d}_i \times \boldsymbol{d}_j\Big) = 0$$

Due to the symmetry of Λ^{ij} the constraint director forces drop out of the last equation. The fact that the constraint director forces (94) do not contribute to the balance of angular momentum can be linked to the rotational invariance of the function $V_\lambda^{\text{RB}} : \mathsf{Q} \rightarrow \mathbb{R}$. This is in complete analogy to Remark 2.3. Due to the arbitrariness of $\boldsymbol{\xi} \in \mathbb{R}^3$, the last equation yields the balance of angular momentum

$$\begin{aligned} \frac{d}{dt}\Big(E_0^{ij}\boldsymbol{d}_i \times \boldsymbol{v}_j\Big) &= \boldsymbol{d}_i \times \boldsymbol{f}^i_{\text{ext}} \\ \frac{d}{dt}\bar{\boldsymbol{j}} &= \overline{\boldsymbol{m}}_{\text{ext}} \end{aligned} \tag{96}$$

relative to the center of mass of the rigid body. Note that the quantities $\bar{\boldsymbol{j}}$ and $\overline{\boldsymbol{m}}_{\text{ext}}$ have been introduced before in (68) and (71), respectively.

3.3 Balance of Energy

Balance of energy can be shown for the rigid body along the lines of the previous treatment of the Cosserat point. Accordingly, scalar multiplying $(95)_2$ by $\boldsymbol{v}_i$ yields

$$\boldsymbol{v}_i \cdot \Big(E_0^{ij}\dot{\boldsymbol{v}}_j - \boldsymbol{f}^i_{\text{ext}} + \Lambda^{ij}\boldsymbol{d}_j\Big) = 0 \tag{97}$$

Due to the symmetry of Λ^{ij} the rate of work done by the constraint director forces (94) can be written as

$$\begin{aligned}\boldsymbol{f}_{\mathrm{c}}^{i} \cdot \boldsymbol{v}_i &= \Lambda^{ij} \boldsymbol{d}_j \cdot \boldsymbol{v}_i \\ &= \Lambda^{ij} \tfrac{1}{2} (\boldsymbol{d}_i \cdot \boldsymbol{v}_j + \boldsymbol{d}_j \cdot \boldsymbol{v}_i) \\ &= 0\end{aligned}$$

The last equality holds due to the constraints on the velocity level (92). This property complies with the fact that ideal forces of constraint are workless. Finally, (97) can be recast in the form

$$\frac{d}{dt}\left(\frac{1}{2} E_0^{ij} \boldsymbol{v}_i \cdot \boldsymbol{v}_j\right) = \boldsymbol{f}_{\mathrm{ext}}^{i} \cdot \boldsymbol{v}_i$$

This is the balance of energy for the rotational motion of the rigid body about its center of mass. The last equation can also be written as $\frac{d}{dt} \overline{T} = \overline{P}_{\mathrm{ext}}$, where $\overline{T}$ denotes the relative kinetic energy introduced in (73) and

$$\overline{P}_{\mathrm{ext}} = \boldsymbol{f}_{\mathrm{ext}}^{i} \cdot \boldsymbol{v}_i \tag{98}$$

denotes the power of the external director forces. Note that the quantities $\overline{T}$ and (98) correspond to the quantities (50) and (48), respectively, in the previous treatment of the Cosserat point.

3.4 Connection with the Classical Euler's Equations

We next link the present equations for the rotational motion of the rigid body to the classical Euler's equations. To this end we recast $(95)_2$ in the form

$$\delta \boldsymbol{d}_i \cdot \left(E_0^{ij} \dot{\boldsymbol{v}}_j - \boldsymbol{f}_{\mathrm{ext}}^{i} + \Lambda^{ij} \boldsymbol{d}_j\right) = 0 \tag{99}$$

which has to hold for arbitrary $\delta \boldsymbol{d}_i \in \mathbb{R}^3$. Now we impose $\delta \boldsymbol{d}_i \in T_{d_i} \mathsf{Q}$. With regard to (91), we set $\delta \boldsymbol{d}_i = \delta \boldsymbol{\vartheta} \times \boldsymbol{d}_i$ for any $\delta \boldsymbol{\vartheta} \in \mathbb{R}^3$ such that (99) can be rewritten as

$$\delta \boldsymbol{\vartheta} \cdot \left(E_0^{ij} \boldsymbol{d}_i \times \dot{\boldsymbol{v}}_j - \boldsymbol{d}_i \times \boldsymbol{f}_{\mathrm{ext}}^{i}\right) = 0$$

Note that in analogy to Sect. 3.2 the constraint director forces drop out of the last equation. The last equation can also be written as

$$E_0^{ij} \boldsymbol{d}_i \times \dot{\boldsymbol{v}}_j = \overline{\boldsymbol{m}}_{\mathrm{ext}} \tag{100}$$

where $\overline{\boldsymbol{m}}_{\text{ext}}$ denotes the resultant external torque relative to the center of mass (see (71)). We next introduce the angular velocity $\boldsymbol{\omega} \in \mathbb{R}^3$ to confine the director velocities to $\boldsymbol{v}_i \in T_{d_i}\mathsf{Q}$. Accordingly, we have $\boldsymbol{v}_i = \boldsymbol{\omega} \times \boldsymbol{d}_i$ such that

$$\begin{aligned}\dot{\boldsymbol{v}}_j &= \dot{\boldsymbol{\omega}} \times \boldsymbol{d}_j + \boldsymbol{\omega} \times \boldsymbol{v}_j \\ &= \dot{\boldsymbol{\omega}} \times \boldsymbol{d}_j + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{d}_j)\end{aligned}$$

Now the left-hand side of (100) can be written as

$$\begin{aligned}E_0^{ij}\boldsymbol{d}_i \times \dot{\boldsymbol{v}}_j &= E_0^{ij}\boldsymbol{d}_i \times (\dot{\boldsymbol{\omega}} \times \boldsymbol{d}_j) + E_0^{ij}\boldsymbol{d}_i \times (\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{d}_j)) \\ &= E_0^{ij}\boldsymbol{d}_i \times (\dot{\boldsymbol{\omega}} \times \boldsymbol{d}_j) + \boldsymbol{\omega} \times (E_0^{ij}\boldsymbol{d}_i \times (\boldsymbol{\omega} \times \boldsymbol{d}_j))\end{aligned} \tag{101}$$

The last equality can be verified by a straightforward calculation using the properties of the vector triple product along with the symmetry of E_0^{ij}. Next consider

$$\begin{aligned}E_0^{ij}\boldsymbol{d}_i \times (\boldsymbol{a} \times \boldsymbol{d}_j) &= E_0^{ij}[(\boldsymbol{d}_i \cdot \boldsymbol{d}_j)\boldsymbol{a} - (\boldsymbol{d}_i \cdot \boldsymbol{a})\boldsymbol{d}_j] \\ &= E_0^{ij}[\delta_{ij}\boldsymbol{I} - \boldsymbol{d}_j \otimes \boldsymbol{d}_i]\boldsymbol{a} \\ &= [\text{tr}(\boldsymbol{E})\boldsymbol{I} - \boldsymbol{E}]\boldsymbol{a} \\ &= \boldsymbol{J}\boldsymbol{a}\end{aligned} \tag{102}$$

for any $\boldsymbol{a} \in \mathbb{R}^3$. Here, the current Euler tensor

$$\boldsymbol{E} = \boldsymbol{F}\boldsymbol{E}_0\boldsymbol{F}^T = E_0^{ij}\boldsymbol{d}_i \otimes \boldsymbol{d}_j$$

has been introduced. Note that $\boldsymbol{E}$ has the same coefficients as the referential Euler tensor (9). Moreover, in (102) the rigid body constraints, namely $d_{ij} = \delta_{ij}$, have been taken into account. Eventually, the classical inertia tensor

$$\boldsymbol{J} = \text{tr}(\boldsymbol{E})\boldsymbol{I} - \boldsymbol{E}$$

has been introduced. Now we are in a position to recast (100) in the form

$$\boldsymbol{J}\dot{\boldsymbol{\omega}} + \boldsymbol{\omega} \times \boldsymbol{J}\boldsymbol{\omega} = \overline{\boldsymbol{m}}_{\text{ext}}$$

which corresponds to the classical Euler's equations for the rigid body.

3.5 EM Integrator for the Rigid Body

As has been shown above, the equations of motion for the rigid body can be directly deduced from those for the hyperelastic Cosserat point by replacing the strain energy (13) with the augmented potential function (87). To construct an EM scheme we apply the notion of a discrete derivative to the new potential function (87). That is, in analogy to the continuous formulation (94), the discrete version of the constraint director forces is given by

$$
\begin{aligned}
(\boldsymbol{f}_{\mathrm{c}}^{i})\big|_{n+\frac{1}{2}} &= \overline{\nabla}_{d_i} V_{\lambda}^{\mathrm{RB}}(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) \\
&= \sum_{l=1}^{6} \lambda^{l} \overline{\nabla}_{d_i} g_l(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) \\
&= 2\mathsf{D}\tilde{V}_{\lambda}^{\mathrm{RB}}(d_{ik_n}, d_{ik_{n+1}})\boldsymbol{d}_{k_{n+\frac{1}{2}}} \\
&= \Lambda^{ik}\boldsymbol{d}_{k_{n+\frac{1}{2}}}
\end{aligned}
\tag{103}
$$

Now the rigid body variant of the EM scheme (80) for the hyperelastic Cosserat point can be written as follows. Given $\boldsymbol{d}_{i_n} \in \mathsf{Q}$ and $\boldsymbol{v}_{i_n} \in \mathbb{R}^3$, $(i = 1, 2, 3)$, find $(\boldsymbol{d}_{i_{n+1}}, \boldsymbol{v}_{i_{n+1}}) \in \mathbb{R}^3 \times \mathbb{R}^3$, and $\lambda_{l_{n+1}} \in \mathbb{R}$ $(l = 1, \ldots, 6)$, as the solution of the algebraic system of equations

$$
\begin{aligned}
\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n} &= \Delta t \boldsymbol{v}_{i_{n+\frac{1}{2}}} \\
E_0^{ij}\left(\boldsymbol{v}_{j_{n+1}} - \boldsymbol{v}_{j_n}\right) &= \Delta t\left((\boldsymbol{f}_{\mathrm{ext}}^{i})\big|_{n+\frac{1}{2}} - \Lambda_{n+1}^{ij}\boldsymbol{d}_{j_{n+\frac{1}{2}}}\right) \\
\hat{g}_l(\boldsymbol{d}_{i_{n+1}}) &= 0
\end{aligned}
\tag{104}
$$

The scheme (104) provides 24 algebraic equations for the determination of the 18 state space coordinates $(\boldsymbol{d}_{i_{n+1}}, \boldsymbol{v}_{i_{n+1}})$ and the 6 independent Lagrange multipliers in Λ_{n+1}^{ij}. We further remark that $(104)_3$ ensures that $\boldsymbol{d}_{i_{n+1}} \in \mathsf{Q}$.

3.5.1 Algorithmic Conservation of Energy

This can be shown as before (see Sect. 2.7.2). One just has to replace the stress resultants $\overline{S}_A^{ij}$ with the Lagrange multipliers Λ_{n+1}^{ij}. Accordingly, combining $(104)_1$ and $(104)_2$ using the dot product yields

$$
\left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \cdot \left((\boldsymbol{f}_{\mathrm{ext}}^{i})\big|_{n+\frac{1}{2}} - \Lambda_{n+1}^{ij}\boldsymbol{d}_{j_{n+\frac{1}{2}}}\right) = \boldsymbol{v}_{i_{n+\frac{1}{2}}} \cdot E_0^{ij}\left(\boldsymbol{v}_{j_{n+1}} - \boldsymbol{v}_{j_n}\right)
$$

or

$$
\Delta t\ (\boldsymbol{f}_{\mathrm{ext}}^{i})\big|_{n+\frac{1}{2}} \cdot \boldsymbol{v}_{i_{n+\frac{1}{2}}} - \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \cdot (\boldsymbol{f}_{\mathrm{c}}^{i})\big|_{n+\frac{1}{2}} = \overline{T}_{n+1} - \overline{T}_n
\tag{105}
$$

On the right-hand side of the last equation the relative kinetic energy $\overline{T}$ (see (73)) has been introduced. On the left-hand side the discrete constraint director forces (103) have been used. Now consider

$$
\begin{aligned}
\left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \cdot (\boldsymbol{f}_{\mathrm{c}}^{i})\big|_{n+\frac{1}{2}} &= \sum_{l=1}^{6} \lambda_{n+1}^{l} \overline{\nabla}_{d_i} g_l(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \\
&= \sum_{l=1}^{6} \lambda_{n+1}^{l}\left(\hat{g}_l(\boldsymbol{d}_{j_{n+1}}) - \hat{g}_l(\boldsymbol{d}_{j_n})\right) \\
&= 0
\end{aligned}
\tag{106}
$$

In the last equation use has been made of the directionality property of the discrete derivative, see (82). In the present context we have

$$\hat{g}_l(\boldsymbol{d}_{j_{n+1}}) - \hat{g}_l(\boldsymbol{d}_{j_n}) = \overline{\nabla}_{d_i} g_l(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \tag{107}$$

Since the EM scheme satisfies the constraints at the end-point of the time step, see (104))$_3$, result (106) follows. Accordingly, in analogy to the continuous case the discrete constraint forces do no work. Altogether, (105) yields the discrete balance equation for the energy

$$\Delta t \ (\overline{P}_{\text{ext}})\big|_{n+\frac{1}{2}} = \overline{T}_{n+1} - \overline{T}_n$$

Accordingly, if the work of the external loading vanishes, the EM scheme conserves the energy.

Remark 3.2 Instead of using the directionality property (107), result (106) can be obtained as well by a direct calculation. To this end consider the work done by the constraint forces in the time interval $[t_n, t_{n+1}]$:

$$\begin{aligned} \Delta t \ (\boldsymbol{f}_{\text{c}}^i)\big|_{n+\frac{1}{2}} \cdot \boldsymbol{v}_{i_{n+\frac{1}{2}}} &= (\boldsymbol{f}_{\text{c}}^i)\big|_{n+\frac{1}{2}} \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \\ &= \Lambda_{n+1}^{ij} \boldsymbol{d}_{j_{n+\frac{1}{2}}} \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \\ &= \Lambda_{n+1}^{ij} \tfrac{1}{2}\left(\boldsymbol{d}_{j_{n+1}} + \boldsymbol{d}_{j_n}\right) \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \\ &= \Lambda_{n+1}^{ij} \tfrac{1}{2}\left(d_{ij_{n+1}} - d_{ij_n}\right) \\ &= 0 \end{aligned} \tag{108}$$

Here, it has been taken into account that (104)$_3$ enforces the algebraic constraints at the end-point of each time step such that $d_{ij_n} = d_{ij_{n+1}} = \delta_{ij}$.

Remark 3.3 Whereas the EM scheme (104) enforces the constraints on the position level explicitly through (104)$_3$, this is not the case for the constraints on the velocity level

$$\frac{d}{dt}\hat{g}_l(\boldsymbol{d}_j) = \frac{\partial \hat{g}_l}{\partial \boldsymbol{d}_j} \cdot \boldsymbol{v}_j = 0 \tag{109}$$

However, due to the directionality property (107) of the discrete derivative applied to the constraints, in the discrete setting the relationship

$$\begin{aligned} \overline{\nabla}_{d_i} g_l(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) \cdot \boldsymbol{v}_{i_{n+\frac{1}{2}}} &= \tfrac{1}{\Delta t}\overline{\nabla}_{d_i} g_l(\boldsymbol{d}_{j_n}, \boldsymbol{d}_{j_{n+1}}) \cdot \left(\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n}\right) \\ &= \hat{g}_l(\boldsymbol{d}_{j_{n+1}}) - \hat{g}_l(\boldsymbol{d}_{j_n}) \\ &= 0 \end{aligned}$$

holds. The last equation can be viewed as discrete counterpart of (109). In analogy to (92), the last equation can be recast in the form

$$\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot \boldsymbol{v}_{j_{n+\frac{1}{2}}} + \boldsymbol{v}_{i_{n+\frac{1}{2}}} \cdot \boldsymbol{d}_{j_{n+\frac{1}{2}}} = 0$$

Accordingly, the rigid body constraints on the velocity level are satisfied at the mid-point of each time step.

3.6 The Director Triad in the Discrete Setting

The present rigid body formulation is based on the canonical embedding of the rotation group SO(3) into the 9-dimensional linear space. Correspondingly, the 3×3 matrix corresponding to the rotation tensor $\boldsymbol{F} \in \mathrm{SO}(3)$ is viewed as vector in $\mathbb{R}^9$ composed of the three directors $\boldsymbol{d}_i \in \mathbb{R}^3$. Due to the present discretization in time, the configuration constraints are relaxed to specific points in time lying on the boundary of the time intervals $[t_n, t_{n+1}]$ $(n = 0, 1, \ldots)$. Accordingly, for finite time steps $\Delta t = t_{n+1} - t_n$, the orthonormality of the director triad $\{\boldsymbol{d}_i\}$ is generally violated inside the time interval $[t_n, t_{n+1}]$. This observation holds in particular for the mid-points $t_{n+\frac{1}{2}} = \frac{1}{2}(t_n + t_{n+1})$.

3.6.1 Planar Rotations

In a first step we investigate the violation of the orthonormality of the mid-point directors for planar rotations. To this end we consider rotations of the rigid body that take place in the plane spanned by the Cartesian base vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. By introducing an angle $\alpha \in \mathbb{R}$, the orthonormality of the director frame can be ensured for arbitrarily large rotation angles (see Fig. 7):

$$\begin{aligned}\widetilde{\boldsymbol{d}}_1(\alpha) &= \cos\alpha \boldsymbol{e}_1 + \sin\alpha \boldsymbol{e}_2\\ \widetilde{\boldsymbol{d}}_2(\alpha) &= -\sin\alpha \boldsymbol{e}_1 + \cos\alpha \boldsymbol{e}_2\end{aligned}$$

and $\boldsymbol{d}_3 = \boldsymbol{e}_3$. Since in the discrete setting the orthonormality condition is always enforced at the endpoints of the time steps we write

$$\begin{aligned}\boldsymbol{d}_{1_n} &= \widetilde{\boldsymbol{d}}_1(\alpha_n)\,, \quad \boldsymbol{d}_{1_{n+1}} = \widetilde{\boldsymbol{d}}_1(\alpha_{n+1})\\ \boldsymbol{d}_{2_n} &= \widetilde{\boldsymbol{d}}_2(\alpha_n)\,, \quad \boldsymbol{d}_{2_{n+1}} = \widetilde{\boldsymbol{d}}_2(\alpha_{n+1})\end{aligned}$$

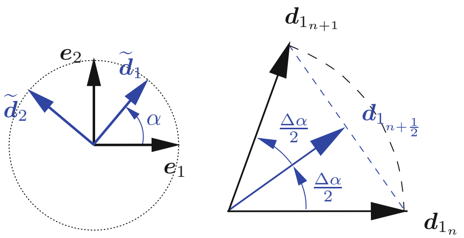

$\boldsymbol{e}_2$
$\widetilde{\boldsymbol{d}}_1$
$\widetilde{\boldsymbol{d}}_2$
α
$\boldsymbol{e}_1$
$\boldsymbol{d}_{1_{n+1}}$
$\frac{\Delta\alpha}{2}$
$\boldsymbol{d}_{1_{n+\frac{1}{2}}}$
$\frac{\Delta\alpha}{2}$
$\boldsymbol{d}_{1_n}$

Fig. 7 *Left* Finite rotation of the director frame about the axis $\boldsymbol{e}_3$ with angle α. *Right* Incremental rotation of $\boldsymbol{d}_1$ with angle $\Delta\alpha$ and corresponding mid-point director $\boldsymbol{d}_{1_{n+\frac{1}{2}}}$

where

$$\alpha_{n+1} = \alpha_n + \Delta\alpha$$

so that the incremental rotation from t_n to t_{n+1} is characterized by the angle $\Delta\alpha$. Now a straightforward calculation shows that the mid-point directors $\boldsymbol{d}_{\beta_{n+\frac{1}{2}}}$ ($\beta = 1, 2$) can be written as

$$\begin{aligned} \boldsymbol{d}_{\beta_{n+\frac{1}{2}}} &= \frac{1}{2}\left(\boldsymbol{d}_{\beta_n} + \boldsymbol{d}_{\beta_{n+1}}\right) \\ &= A(\Delta\alpha)\widetilde{\boldsymbol{d}}_\beta(\alpha_{n+\frac{1}{2}}) \end{aligned}$$

where $\alpha_{n+\frac{1}{2}} = \frac{1}{2}(\alpha_n + \alpha_{n+1})$, and

$$A(\Delta\alpha) = \cos\left(\frac{\Delta\alpha}{2}\right)$$

Accordingly, the mid-point approximation of the directors is still orthogonal, for

$$\boldsymbol{d}_{1_{n+\frac{1}{2}}} \cdot \boldsymbol{d}_{2_{n+\frac{1}{2}}} = \left(A(\Delta\alpha)\right)^2 \widetilde{\boldsymbol{d}}_1(\alpha_{n+\frac{1}{2}}) \cdot \widetilde{\boldsymbol{d}}_2(\alpha_{n+\frac{1}{2}}) = 0$$

but generally fails to be of unit length (see also Fig. 7). In particular, we have

$$\begin{aligned} \boldsymbol{d}_{(\beta)_{n+\frac{1}{2}}} \cdot \boldsymbol{d}_{(\beta)_{n+\frac{1}{2}}} &= \left(A(\Delta\alpha)\right)^2 \parallel \widetilde{\boldsymbol{d}}_\beta(\alpha_{n+\frac{1}{2}}) \parallel^2 \\ &= \tfrac{1}{2}\left(1 + \cos(\Delta\alpha)\right) \\ &\leq 1 \end{aligned} \tag{110}$$

To summarize, in the case of planar rotations, the mid-point directors stay mutually orthogonal but their length is reduced. Note that this discretization error decreases if the rotation increment (or time step) is reduced.

3.6.2 Three-dimensional Rotations

In the three-dimensional setting the mid-point directors are in general neither of unit length, nor mutually orthogonal. That is, $\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot \boldsymbol{d}_{j_{n+\frac{1}{2}}} \neq \delta_{ij}$ in general. In particular, a lengthy but straightforward calculation, employing the well-known formula (Bottema and Roth 1979; Hughes and Winget 1980)

$$\boldsymbol{d}_{i_{n+1}} - \boldsymbol{d}_{i_n} = \boldsymbol{\vartheta} \times \boldsymbol{d}_{i_{n+\frac{1}{2}}}$$

shows that

$$\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot \boldsymbol{B}(\boldsymbol{\vartheta})\boldsymbol{d}_{j_{n+\frac{1}{2}}} = \delta_{ij} \tag{111}$$

where

$$\boldsymbol{B}(\boldsymbol{\vartheta}) = \left(1 + \frac{1}{4}\boldsymbol{\vartheta} \cdot \boldsymbol{\vartheta}\right)\boldsymbol{I} - \frac{1}{4}\boldsymbol{\vartheta} \otimes \boldsymbol{\vartheta}$$

Thus, in the limit case of vanishing incremental rotations (i.e., $\boldsymbol{\vartheta} = \mathbf{0}$) we get $\boldsymbol{B}(\mathbf{0}) = \boldsymbol{I}$, and the orthonormality of the mid-point directors is recovered.

Moreover, if $\boldsymbol{\vartheta} \cdot \boldsymbol{d}_{i_{n+\frac{1}{2}}} = 0$, such as in the planar case, Eq. (111) yields

$$\left(1 + \frac{1}{4}\boldsymbol{\vartheta} \cdot \boldsymbol{\vartheta}\right)\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot \boldsymbol{d}_{j_{n+\frac{1}{2}}} = \delta_{ij}$$

Accordingly, the mid-point directors are mutually orthogonal. In addition to that, the relation $\boldsymbol{d}_{(i)_{n+\frac{1}{2}}} \cdot \boldsymbol{d}_{(i)_{n+\frac{1}{2}}} = (1 + \frac{1}{4}\boldsymbol{\vartheta} \cdot \boldsymbol{\vartheta})^{-1}$ shows that the length of the mid-point directors is generally smaller than one. This result is in agreement with (110). In particular, it can be shown that

$$\begin{aligned}(A(\Delta\alpha))^2 &= \tfrac{1}{2}\left(1 + \cos(\Delta\alpha)\right)\\ &= \left(1 + \left(\tfrac{\|\boldsymbol{\vartheta}\|}{2}\right)^2\right)^{-1}\end{aligned}$$

for $\frac{\|\boldsymbol{\vartheta}\|}{2} = \tan(\frac{\Delta\alpha}{2})$.

Remark 3.4 Similar geometric considerations apply to the elastic Cosserat point. In particular, the application of the mid-point rule rests on the Green–Lagrangean strains

$$\gamma_{ij}^{\mathrm{MP}} = \frac{1}{2}\left(d_{ij}^{\mathrm{MP}} - \delta_{ij}\right) \tag{112}$$

where

$$d_{ij}^{\mathrm{MP}} = \boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot \boldsymbol{d}_{j_{n+\frac{1}{2}}}$$

Accordingly, if the elastic Cosserat point undergoes finite rotations, the mid-point rule in general generates artificial strains. This discretization error is especially pronounced for stiff material behavior and might trigger spurious oscillations leading to numerical instabilities. Originally, artificial normal strains produced by the mid-point rule have been observed in the context of an elastic pendulum (Tarnow 1993; Crisfield and Shi 1994).

3.7 The Link to Natural Coordinates

The present formulation of rigid body dynamics is closely related to the notion of natural coordinates advocated by García de Jalón and co-workers (García de Jalón

2007). This can be easily shown by considering the connection between the present coordinates and the natural coordinates associated with the *most general element* (García de Jalón and Bayo 1994). The configuration of the most general element is specified by

$$\boldsymbol{q}^e = \left[\boldsymbol{r}_A^T \; \boldsymbol{r}_B^T \; \boldsymbol{u}^T \;\; \boldsymbol{v}^T\right]^T \tag{113}$$

where $\boldsymbol{r}_A, \boldsymbol{r}_B \in \mathbb{R}^3$ denote the position vectors of two basic points A, B, and $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$ denote two non-coplanar unit vectors (Fig. 8). The natural coordinates in (113) can now be expressed in terms of the present coordinates:

$$\begin{aligned} \boldsymbol{r}_A &= \bar{\boldsymbol{x}} + X_A^i \boldsymbol{d}_i \\ \boldsymbol{r}_B &= \bar{\boldsymbol{x}} + X_B^i \boldsymbol{d}_i \end{aligned} \qquad \text{and} \qquad \begin{aligned} \boldsymbol{u} &= U^i \boldsymbol{d}_i \\ \boldsymbol{v} &= V^i \boldsymbol{d}_i \end{aligned}$$

Here X_A^i, X_B^i are the material coordinates of points A, B, and U^i, V^i are the components of the unit vectors $\boldsymbol{u}, \boldsymbol{v}$ relative to the director (or body) frame. Alternatively, we may write

$$\boldsymbol{q}^e = \boldsymbol{T}\boldsymbol{q}$$

where $\boldsymbol{q} \in \mathbb{R}^{12}$ contains the present coordinates according to (58), and $\boldsymbol{T}$ is a 12×12 transformation matrix of the form

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{I} & X_A^1\boldsymbol{I} & X_A^2\boldsymbol{I} & X_A^3\boldsymbol{I} \\ \boldsymbol{I} & X_B^1\boldsymbol{I} & X_B^2\boldsymbol{I} & X_B^3\boldsymbol{I} \\ \boldsymbol{0} & U^1\boldsymbol{I} & U^2\boldsymbol{I} & U^3\boldsymbol{I} \\ \boldsymbol{0} & V^1\boldsymbol{I} & V^2\boldsymbol{I} & V^3\boldsymbol{I} \end{bmatrix}$$

The mass matrix pertaining to the most general element is given by

$$\boldsymbol{M}^e = \boldsymbol{T}^T\boldsymbol{M}\boldsymbol{T}$$

where the constant mass matrix $\boldsymbol{M}$ of the present formulation is given by (60). Since $\boldsymbol{T}$ is *constant*, $\boldsymbol{M}^e$ is constant too. The connection between further rigid body elements

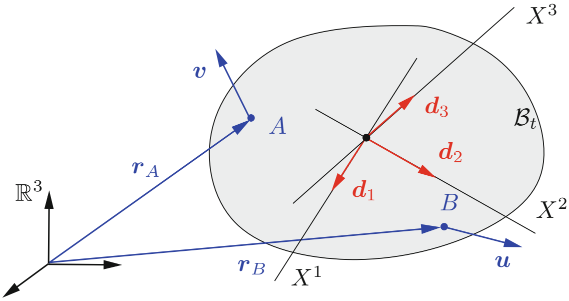


Fig. 8 Connection between the present director formulation and natural coordinates

belonging to the family of elements provided by the natural coordinates approach can be found in García de Jalón and Bayo (1994, Sect. 4.2.2).

3.8 Application of External Torques

The application of external torques $\overline{\boldsymbol{m}}_{\text{ext}}$ relative to the center of mass of the rigid body can be accomplished via the external director forces $\boldsymbol{f}^i_{\text{ext}}$, cf. (71). For this purpose one may use

$$\begin{bmatrix} \boldsymbol{f}^1_{\text{ext}} \\ \boldsymbol{f}^2_{\text{ext}} \\ \boldsymbol{f}^3_{\text{ext}} \end{bmatrix} = \frac{1}{2\sqrt{d}} \begin{bmatrix} \boldsymbol{d}_2 \otimes \boldsymbol{d}_3 - \boldsymbol{d}_3 \otimes \boldsymbol{d}_2 \\ \boldsymbol{d}_3 \otimes \boldsymbol{d}_1 - \boldsymbol{d}_1 \otimes \boldsymbol{d}_3 \\ \boldsymbol{d}_1 \otimes \boldsymbol{d}_2 - \boldsymbol{d}_2 \otimes \boldsymbol{d}_1 \end{bmatrix} \overline{\boldsymbol{m}}_{\text{ext}} = \frac{1}{2} \begin{bmatrix} \overline{\boldsymbol{m}}_{\text{ext}} \times \boldsymbol{d}^1 \\ \overline{\boldsymbol{m}}_{\text{ext}} \times \boldsymbol{d}^2 \\ \overline{\boldsymbol{m}}_{\text{ext}} \times \boldsymbol{d}^3 \end{bmatrix}$$

This relationship is derived in Appendix A.2. In the discrete setting we make use of

$$(\boldsymbol{f}^i_{\text{ext}})\Big|_{n+\frac{1}{2}} = \frac{1}{2}\, \overline{\boldsymbol{m}}_{\text{ext}}|_{n+\frac{1}{2}} \times \boldsymbol{d}^i\Big|_{n+\frac{1}{2}} \tag{114}$$

Here, $\overline{\boldsymbol{m}}_{\text{ext}}|_{n+\frac{1}{2}}$ represents an external torque applied in the time interval $[t_n, t_{n+1}]$, and $\boldsymbol{d}^i\Big|_{n+\frac{1}{2}}$ are contravariant mid-point directors that satisfy the condition

$$\boldsymbol{d}^i\Big|_{n+\frac{1}{2}} \cdot \boldsymbol{d}_{j_{n+\frac{1}{2}}} = \delta^i_j$$

To satisfy the balance of angular momentum in the discrete setting, it is of paramount importance to distinguish between covariant mid-point directors, $\boldsymbol{d}_{j_{n+\frac{1}{2}}}$, and associated contravariant (or dual) mid-point directors, $\boldsymbol{d}^i\Big|_{n+\frac{1}{2}}$. This fact is closely related to the properties of the mid-point directors investigated in Sect. 3.6. Formula (114) has originally been proposed in Betsch et al. (2012), see also Betsch and Sänger (2013) and Koch and Leyendecker (2013).

3.9 Balance of Angular Momentum in the Discrete Setting

We next prove that formula (114) does indeed make possible the consistent application of external torques. To this end we consider the discrete counterpart of the continuous relationship $\frac{d}{dt}\bar{\boldsymbol{j}} = \overline{\boldsymbol{m}}_{\text{ext}}$, see $(96)_2$, which is given by

$$\bar{\boldsymbol{j}}_{n+1} - \bar{\boldsymbol{j}}_n = \Delta t \boldsymbol{d}_{j_{n+\frac{1}{2}}} \times (\boldsymbol{f}^j_{\text{ext}})\Big|_{n+\frac{1}{2}}$$

cf. (70). Inserting from (114) yields

$$
\begin{aligned}
\bar{\boldsymbol{j}}_{n+1} - \bar{\boldsymbol{j}}_n &= \frac{\Delta t}{2} \boldsymbol{d}_{i_{n+\frac{1}{2}}} \times \left(\overline{\boldsymbol{m}}_{\mathrm{ext}}|_{n+\frac{1}{2}} \times \boldsymbol{d}^i\big|_{n+\frac{1}{2}} \right) \\
&= \frac{\Delta t}{2} \left((\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot \boldsymbol{d}^i\big|_{n+\frac{1}{2}}) \, \overline{\boldsymbol{m}}_{\mathrm{ext}}|_{n+\frac{1}{2}} - (\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot \overline{\boldsymbol{m}}_{\mathrm{ext}}|_{n+\frac{1}{2}}) \, \boldsymbol{d}^i\big|_{n+\frac{1}{2}} \right) \\
&= \Delta t \, \overline{\boldsymbol{m}}_{\mathrm{ext}}|_{n+\frac{1}{2}}
\end{aligned}
\tag{115}
$$

Consequently, formula (114) guarantees that external torques are properly applied in the discrete setting.

4 Extension to Multibody Dynamics

So far we focused on a single Cosserat point and a single rigid body. However, the present framework can be easily extended to nonlinear structural dynamics and flexible multibody dynamics by applying Cosserat theories for the description of nonlinear beams and shells (Rubin 2000; Antman 2005; Bauchau 2011). Further details of the extension of the present approach to more complicated mechanical systems may be found in Betsch and Steinmann (2002b, c, 2003), Betsch (2006), Betsch and Leyendecker (2006), Leyendecker et al. (2006, 2008a), Betsch and Uhlar (2007), Betsch and Sänger (2009a, b).

In this work we illustrate the extension of the present approach to classical multibody systems, comprised of rigid bodies. First we consider the formulation of kinematic pairs.

4.1 Kinematic Pairs

We next illustrate the formulation of kinematic pairs with the example of a cylindrical pair (Fig. 9). To this end we consider two rigid bodies formulated as constrained mechanical systems as described in Sect. 3. Accordingly, the configuration of the two-body system under consideration is characterized by redundant coordinates

$$
\boldsymbol{q} = \begin{bmatrix} {}^1\boldsymbol{q} \\ {}^2\boldsymbol{q} \end{bmatrix} \quad \text{where} \quad {}^\alpha\boldsymbol{q} = \begin{bmatrix} {}^\alpha\boldsymbol{\varphi} \\ {}^\alpha\boldsymbol{d}_1 \\ {}^\alpha\boldsymbol{d}_2 \\ {}^\alpha\boldsymbol{d}_3 \end{bmatrix}
\tag{116}
$$

Note that the contribution of body α to the configuration vector coincides with (58). The equations of motion pertaining to the constrained mechanical system at hand can again be formulated as outlined in Sect. 3. Similar to (116), the contribution of

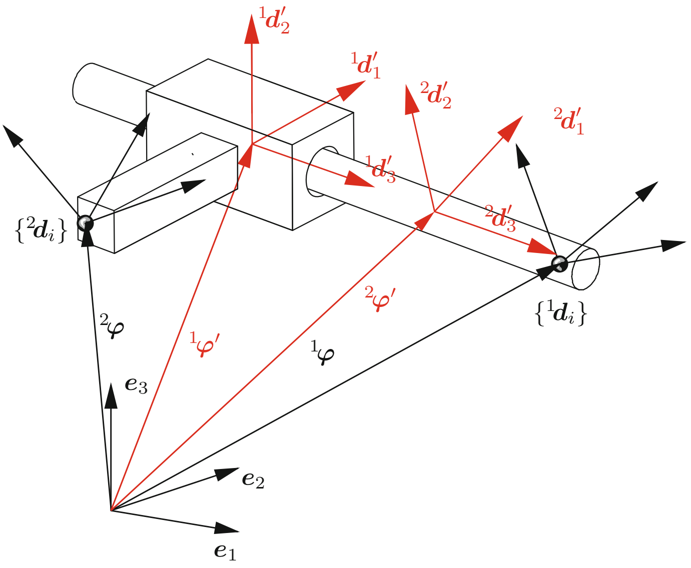


Fig. 9 Sketch of the cylindrical pair: Coordinates (${}^\alpha\boldsymbol{\varphi}$, $\{{}^\alpha\boldsymbol{d}_i\}$) characterizing the current configuration ${}^\alpha\mathcal{B}_t$ of rigid body α. The additional systems (${}^\alpha\boldsymbol{\varphi}'$, $\{{}^\alpha\boldsymbol{d}_i'\}$) are introduced for the description of the motion of the second body relative to the first body (translation along and rotation about ${}^1\boldsymbol{d}_3' = {}^2\boldsymbol{d}_3'$). The connection between (${}^\alpha\boldsymbol{\varphi}'$, $\{{}^\alpha\boldsymbol{d}_i'\}$) and the coordinates (${}^\alpha\boldsymbol{\varphi}$, $\{{}^\alpha\boldsymbol{d}_i\}$) is defined in the initial configuration of the multibody system

each rigid body to the external forces leads to the system vector

$$\boldsymbol{F} = \begin{bmatrix} {}^1\boldsymbol{F} \\ {}^2\boldsymbol{F} \end{bmatrix} \quad \text{where} \quad {}^\alpha\boldsymbol{F} = \begin{bmatrix} {}^\alpha\boldsymbol{f}_\varphi \\ {}^\alpha\boldsymbol{f}^1 \\ {}^\alpha\boldsymbol{f}^2 \\ {}^\alpha\boldsymbol{f}^3 \end{bmatrix} \tag{117}$$

Note that the force vector ${}^\alpha\boldsymbol{F}$ associated with body α coincides with (61).

4.1.1 Initialization of Kinematic Relationships

To describe the motion of the second body relative to the first one we introduce orthonormal body-fixed triads $\{{}^\alpha\boldsymbol{d}_i'\}$ in such a way that the unit vectors ${}^\alpha\boldsymbol{d}_3'$ are parallel to the axis of the cylindrical pair (Fig. 9). Moreover, we choose the two orthonormal triads to coincide in the initial configuration, i.e. ${}^1\boldsymbol{d}_i'(0) = {}^2\boldsymbol{d}_i'(0)$. The connection between the newly introduced orthonormal triads $\{{}^\alpha\boldsymbol{d}_i'\}$ and the original triads $\{{}^\alpha\boldsymbol{d}_i\}$ is given by

$${}^\alpha\boldsymbol{R}' = {}^\alpha\boldsymbol{F}\,{}^\alpha\boldsymbol{\Lambda}_0 \tag{118}$$

where

$${}^{\alpha}\boldsymbol{F} = {}^{\alpha}\boldsymbol{d}_i \otimes \boldsymbol{e}^i \quad \text{and} \quad {}^{\alpha}\boldsymbol{R}' = {}^{\alpha}\boldsymbol{d}'_i \otimes \boldsymbol{e}^i$$

The constant tensors ${}^{\alpha}\boldsymbol{\Lambda}_0$ in (118) are calculated in the initial configuration via

$${}^{\alpha}\boldsymbol{\Lambda}_0 = {}^{\alpha}\boldsymbol{F}^{-1}(0)\, {}^{\alpha}\boldsymbol{R}'(0)$$

The origin of the newly introduced orthonormal triads $\{{}^{\alpha}\boldsymbol{d}'_i\}$ is fixed at material points ${}^{\alpha}\Theta^i$ whose placement in the current configuration ${}^{\alpha}\mathcal{B}_t$ of rigid body α is denoted by ${}^{\alpha}\boldsymbol{\varphi}'$. Accordingly,

$${}^{\alpha}\boldsymbol{\varphi}' = {}^{\alpha}\boldsymbol{\varphi} + {}^{\alpha}\Theta^i\, {}^{\alpha}\boldsymbol{d}_i$$

Note that the location of the material points ${}^{\alpha}\Theta^i$ has to be specified during initialization.

4.1.2 Configuration Space of the Cylindrical Pair

The configuration space of the cylindrical pair can be easily defined by distinguishing between internal constraints due the assumption of rigidity and external constraints due to the interconnection between the rigid bodies in a multibody system (Betsch and Steinmann 2002c). Accordingly, the present description of the cylindrical pair relies on $n = 24$ coordinates subject to 12 internal constraints $\boldsymbol{g}^{\text{int}}({}^{\alpha}\boldsymbol{q}) = \boldsymbol{0}\,(\alpha = 1, 2)$, where $\boldsymbol{g}^{\text{int}} : \mathbb{R}^{12} \rightarrow \mathbb{R}^{6}$ follows from (89), and 4 external constraints associated with the constraint functions

$$\boldsymbol{g}_{\text{P}}^{\text{ext}}(\boldsymbol{q}) = \begin{bmatrix} {}^{1}\boldsymbol{d}'_1 \cdot \left({}^{2}\boldsymbol{\varphi}' - {}^{1}\boldsymbol{\varphi}'\right) \\ {}^{1}\boldsymbol{d}'_2 \cdot \left({}^{2}\boldsymbol{\varphi}' - {}^{1}\boldsymbol{\varphi}'\right) \end{bmatrix} \tag{119}$$

and

$$\boldsymbol{g}_{\text{R}}^{\text{ext}}(\boldsymbol{q}) = \begin{bmatrix} {}^{1}\boldsymbol{d}'_1 \cdot {}^{2}\boldsymbol{d}'_3 \\ {}^{1}\boldsymbol{d}'_2 \cdot {}^{2}\boldsymbol{d}'_3 \end{bmatrix} \tag{120}$$

To summarize, we have $n = 24$ coordinates subject to $m = 16$ constraints which can be assembled in the constraint function $\boldsymbol{g}^{\text{C}} : \mathbb{R}^{24} \rightarrow \mathbb{R}^{16}$ given by

$$\boldsymbol{g}^{\text{C}}(\boldsymbol{q}) = \begin{bmatrix} \boldsymbol{g}^{\text{int}}({}^{1}\boldsymbol{q}) \\ \boldsymbol{g}^{\text{int}}({}^{2}\boldsymbol{q}) \\ \boldsymbol{g}_{\text{P}}^{\text{ext}}(\boldsymbol{q}) \\ \boldsymbol{g}_{\text{R}}^{\text{ext}}(\boldsymbol{q}) \end{bmatrix} \tag{121}$$

Consequently, the configuration space of the cylindrical pair is defined by

$$\mathsf{Q}^{\text{C}} = \{\boldsymbol{q} \in \mathbb{R}^{24} \,|\, \boldsymbol{g}^{\text{C}}(\boldsymbol{q}) = \boldsymbol{0}\} \tag{122}$$

4.2 Multibody Systems

As mentioned before, geometrically exact Cosserat models for beams and shells fit perfectly well into the present framework. In particular, if the nonlinear beam and shell formulations are discretized in space as proposed in Betsch and Steinmann (2002b, 2003), Betsch and Sänger (2009a), the equations of motion pertaining to the resulting discrete mechanical systems fit into the framework outlined in Sect. 3. Thus the use of director coordinates makes possible a uniform formulation of flexible multibody dynamics.[2] Main characteristics of the present approach can be summarized as follows:

1. The inertia parameters are always constant leading to the simple structure of the inertia terms in the equations of motion. In particular, the differential part of the equations of motion can be written as

$$\boldsymbol{M}\ddot{\boldsymbol{q}} + \nabla V_\lambda(\boldsymbol{q}) - \boldsymbol{F} = \boldsymbol{0}$$

   where the potential forces along with the constraint forces can be derived from an augmented potential function of the form

$$V_\lambda(\boldsymbol{q}) = U(\boldsymbol{q}) + \sum_{l=1}^{m} \lambda^l \nabla g_l(\boldsymbol{q})$$

   For example, the potential function $U(\boldsymbol{q})$ can be associated with the action of gravitational forces or with the deformation of flexible bodies such as nonlinear beams and shells relying on hyperelastic constitutive laws.
2. The configuration vector of the complete flexible multibody systems is composed of vectors $\boldsymbol{q}_I \in \mathbb{R}^3$ and thus given by

$$\boldsymbol{q} = \begin{bmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q}_N \end{bmatrix} \tag{123}$$

   where N denotes the total number of 3-vectors $\boldsymbol{q}_I$ needed to describe a specific multibody system. Accordingly, in total, the configuration vector $\boldsymbol{q} \in \mathbb{R}^n$ has $n = 3N$ components.
3. The total angular momentum of flexible multibody systems can be cast in the form

$$\boldsymbol{J} = \sum_{a,b=1}^{N} M^{ab} \boldsymbol{q}_a \times \boldsymbol{v}_b \tag{124}$$

[2] The present framework comprises as well domain decomposition problems (Hesch and Betsch 2010) and large deformation contact (Hesch and Betsch 2009, 2011a, b).

where M^{ab} contain the constant inertia parameters and $\boldsymbol{v}_b = \dot{\boldsymbol{q}}_b$.

4. The balance of angular momentum can be written as

$$\frac{d}{dt}\boldsymbol{J} = \sum_{a=1}^{N} \boldsymbol{q}_a \times \big(\boldsymbol{F}^a - \nabla_{\boldsymbol{q}_a} V_\lambda(\boldsymbol{q})\big) \tag{125}$$

The EM consistent discretization of the discrete mechanical systems at hand can be performed in complete analogy to the Cosserat point and the rigid body dealt with in detail in the previous sections.

5 Numerical Examples

5.1 *Spacecraft Attitude Maneuver*

In the first numerical example we demonstrate the importance of formula (114) for the consistent application of external torques. To this end we apply the present approach to the control of spacecraft rotational maneuvers.

The spacecraft is modeled as multibody system consisting of four rigid bodies (Fig. 10), namely the base body and three reaction wheels. A similar example has been dealt with in Leyendecker et al. (2010). The data for the present 4-body system

Fig. 10 The spacecraft as 4-body system

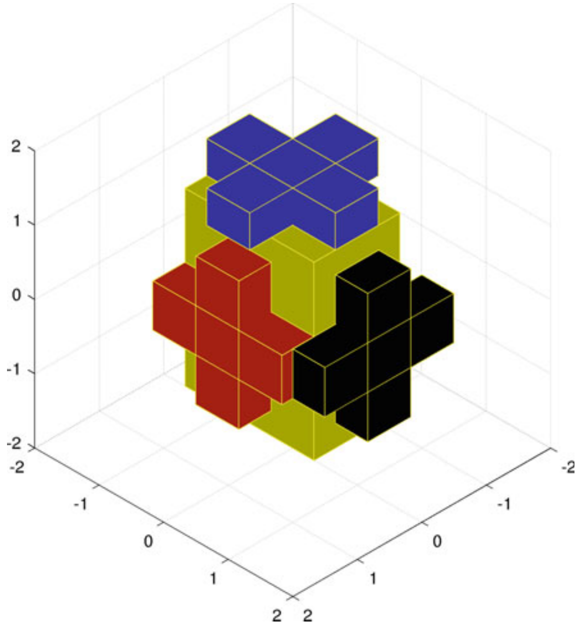


Table 1 Spacecraft: data for the 4-body system

Body	M_φ	E^{11}	E^{22}	E^{33}	L
1	1005.3096	89.3609	201.0619	357.4434	
2	424.1150	8.8357	106.0288	106.0288	0.9167
3	424.1150	106.0288	8.8357	106.0288	1.25
4	424.1150	106.0288	106.0288	8.8357	1.5833

Note that L denotes the distance between the center of mass of the reaction wheels and the base body

have been taken from (Leyendecker et al. 2010). Using principal axis for each rigid body the data used in the simulations are summarized in Table 1.

The reaction wheels are spinning about body-fixed axis of the base body. For simplicity the three body-fixed axis are assumed to coincide with the director frame $\{{}^{1}\boldsymbol{d}_i\}$ of the base body. Spacecraft attitude maneuvers are performed by applying reaction wheel motor torques

$$ {}^{2}\boldsymbol{m} = (u^1)\,{}^{1}\boldsymbol{d}_1\,, \quad {}^{3}\boldsymbol{m} = (u^2)\,{}^{1}\boldsymbol{d}_2\,, \quad {}^{4}\boldsymbol{m} = (u^3)\,{}^{1}\boldsymbol{d}_3 \tag{126} $$

In the example we prescribe constant motor torques $u^i = 200$.

A total of $n = 48$ coordinates are employed to describe the multibody system at hand. Each body is subject to 6 rigid body constraints giving rise to $m^{\text{int}} = 24$ internal constraints. Revolute joints are used to connect the reaction wheels to the base body. This amounts to $m^{\text{ext}} = 3 \times 5 = 15$ external constraints. Accordingly, in total there are $m = m^{\text{int}} + m^{\text{ext}} = 39$ independent constraints leading to $n - m = 9$ degrees of freedom.

The newly devised formula (114) has been used to consistently apply the motor torques to the reaction wheels. The torque acting on the base body is given by

$$ {}^{1}\boldsymbol{m} = -\left({}^{2}\boldsymbol{m} + {}^{3}\boldsymbol{m} + {}^{4}\boldsymbol{m}\right) \tag{127} $$

Since no resultant external torque acts on the spacecraft, the total angular momentum is a first integral of the motion. In particular,

$$ \begin{aligned} \boldsymbol{J}_{n+1} - \boldsymbol{J}_n &= \Delta t \sum_{b=1}^{4} {}^{b}\boldsymbol{d}_{i_{n+\frac{1}{2}}} \times {}^{b}\boldsymbol{f}^{i}\big|_{n+\frac{1}{2}} \\ &= \frac{\Delta t}{2} \sum_{b=1}^{4} {}^{b}\boldsymbol{d}_{i_{n+\frac{1}{2}}} \times \left({}^{b}\boldsymbol{m} \times {}^{b}\boldsymbol{d}^{i}\big|_{n+\frac{1}{2}}\right) \\ &= \frac{\Delta t}{2} \sum_{b=1}^{4} \left(({}^{b}\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot {}^{b}\boldsymbol{d}^{i})\,{}^{b}\boldsymbol{m} - ({}^{b}\boldsymbol{d}_{i_{n+\frac{1}{2}}} \cdot {}^{b}\boldsymbol{m})\,{}^{b}\boldsymbol{d}^{i}\big|_{n+\frac{1}{2}}\right) \\ &= \Delta t \sum_{b=1}^{4} {}^{b}\boldsymbol{m} \\ &= \boldsymbol{0} \end{aligned} $$

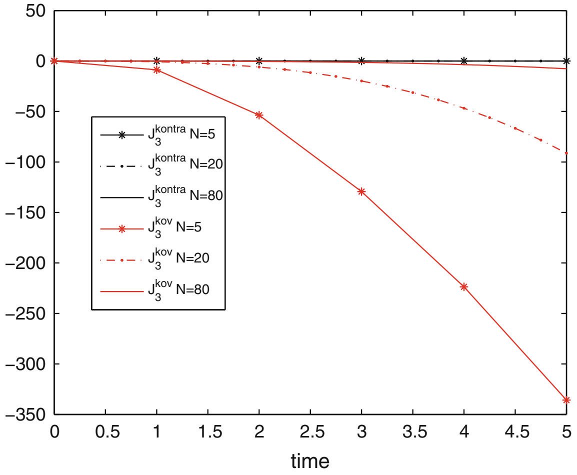


Fig. 11 Spacecraft: Comparison of angular momentum

where use has been made of (126) and (127). In the numerical simulations we focus on the 3-component J_3 of the total angular momentum and the total kinetic energy T of the multibody system at hand. The numerical results due to the application of the newly devised formula (114) are denoted by J_3^{kontra} and T^{kontra}.

For comparison we apply the motor torques via the straightforward mid-point evaluation of the continuous expression of the 'original' formulation (Betsch et al. 2012).

$$\boldsymbol{f}_{i_{n+\frac{1}{2}}} = \frac{1}{2}\boldsymbol{m}_{n+\frac{1}{2}} \times \boldsymbol{d}_{i_{n+\frac{1}{2}}} \tag{128}$$

The corresponding results are denoted by J_3^{kov} and T^{kov}.

A number of N time steps is used to resolve the time interval $[0, 5]$. It can be observed from Fig. 11 that J_3^{kontra} stays constant for all N. This corroborates algorithmic conservation of the total angular momentum. In severe contrast to that J_3^{kov} does not stay constant. Accordingly the balance law for angular momentum is violated. This discretization error can be decreased by raising the number of time steps N. These observations are further supported by considering the total kinetic energy in Fig. 12. Accordingly, T^{kontra} does hardly change if the time steps are refined. That is, using only $N = 5$ time steps already leads to a very good approximation of the kinetic energy. This is in severe contrast to T^{kov}.

5.2 *Parallel Robot*

In the second example we consider the planar parallel robot depicted in Fig. 13. Each of the three legs of the parallel robot consists of a prismatic kinematic pair along with two revolute joints. The parallel mechanism has three degrees of freedom and

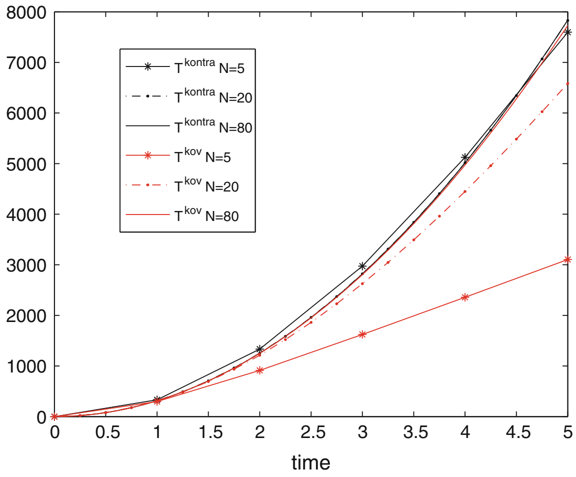


Fig. 12 Spacecraft: Comparison of kinetic energy

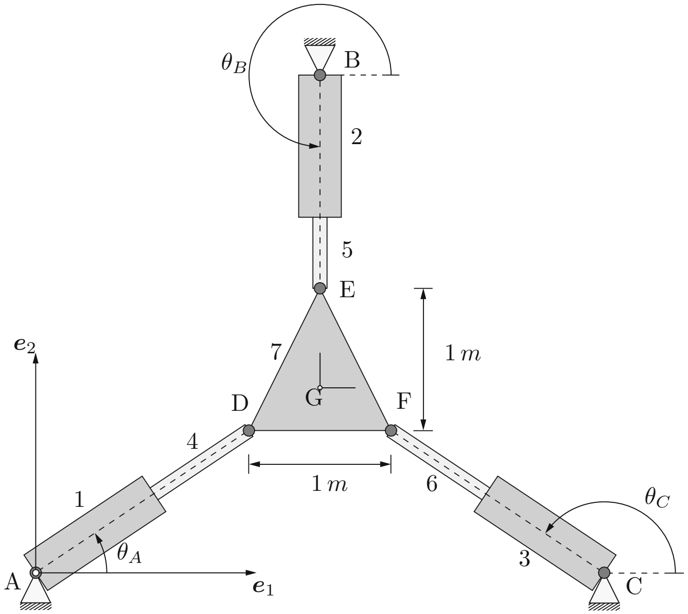


Fig. 13 The 3-<u>R</u>PR planar parallel robot

is referred to as the 3-<u>R</u>PR planar parallel manipulator, where the underlined letter indicates that one of the revolute joints of each leg is driven.

In the forward dynamics simulation we rely on the results of an inverse dynamics analysis due to McPhee and Redmond (2006). The goal of the inverse dynamics analysis is to determine the driving torques required to translate the center of mass G of the end-effector in a figure-8 pattern, with a cycle time of 2 s, defined by

$$\begin{aligned} x_G &= 2 + \sin(\pi t) \\ y_G &= \frac{4}{3} + \frac{1}{2}\sin(2\pi t) \\ \theta_G &= 0 \end{aligned} \tag{129}$$

The geometry and inertia properties of the parallel robot have been taken as well from McPhee and Redmond (2006) and are summarized in Table 2. In addition to that, we remark that the position of points B and C (Fig. 13) is given by $x_B = 2$, $y_B = 3.5$, and $x_C = 4.0$. The result of the inverse dynamics analysis gives rise to the three driving torques, one of which is depicted in Fig. 14 (compare with Fig. 12 in McPhee and Redmond 2006).

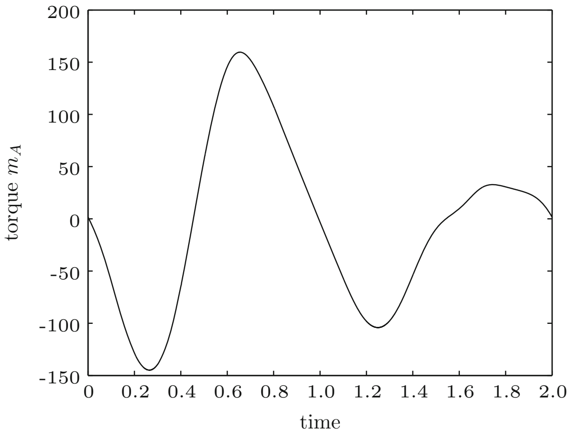
Obviously, using the three driving torques from the inverse dynamics analysis in the forward dynamics simulation along with the data in Table 2 should lead to the motion of the end-effector given by (129). That is, the trajectory of the center of mass G of the end-effector should follow a figure-8 pattern, while the end-effector should not rotate.

In the simulation we use 200 time steps and apply formula (114) for the consistent application of external torques. It can be observed from Fig. 15 that the proposed

Table 2 Geometry and inertia properties of the parallel robot

Body	Width (m)	Length (m)	Mass (kg)	Moment of inertia (kg m^2)
1, 2, 3	0.3	1.0	2.4	0.218
4, 5, 6	0.1	1.5	1.2	0.226
7	1.0	1.0	0.5	0.049

Fig. 14 Parallel robot: Driving torque at joint A determined by the inverse dynamics analysis



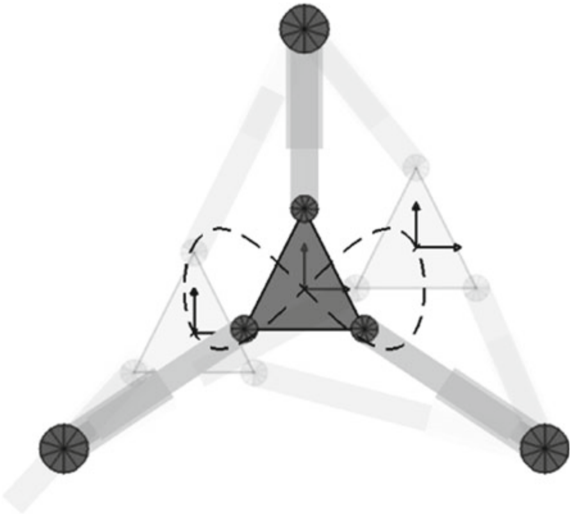


Fig. 15 Parallel robot: Final position simulated with the proposed method. The figure-8 trajectory is correctly tracked by the mass center of the end-effector

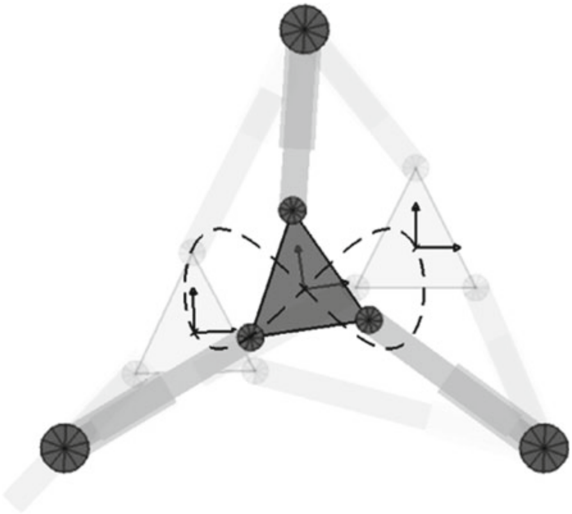


Fig. 16 Parallel robot: Final position simulated with the original method. The inconsistent application of the driving torques leads to a deviation from the correct motion

simulation method yields the correct motion. In sharp contrast to that, using instead of formula (114) the mid-point evaluation of the original formulation, Eq. (128), yields a deviation from the correct motion (Fig. 16). This observation is further supported by Fig. 17, where the rotation angle of the end-effector is plotted versus time. While the advocated method correctly reproduces the constant angle $\theta_G^{kontra} = 0$, the angle θ_G^{kov} determined by the original approach deviates significantly from the correct value. These results strongly support the need for a consistent formulation of external torques in the underlying rotationless formulation.

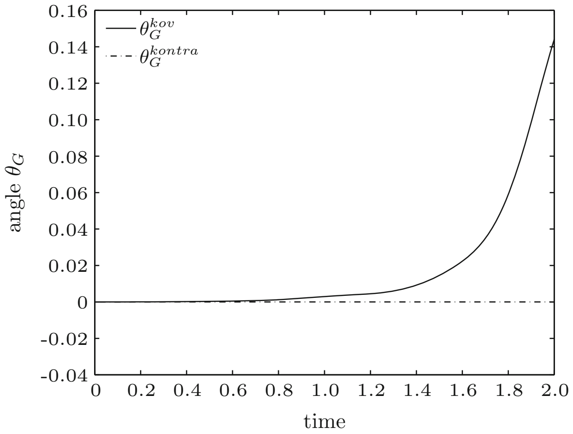


Fig. 17 Parallel robot: Rotation angle of the end-effector

Acknowledgments Support for this research was provided by the Deutsche Forschungsgemeinschaft (DFG) under Grant BE 2285/10-1. This support is gratefully acknowledged.

A Appendix

A.1 Balance Laws

For comparison, the balance laws are directly derived from the variational equations (19) governing the motion of the pseudo-rigid body. To this end, we recast (19) in the form

$$\delta\bar{\boldsymbol{x}} \cdot \left(M\ddot{\bar{\boldsymbol{x}}} - \boldsymbol{f}_{\text{ext}}\right) = 0 \tag{130}$$

$$\operatorname{tr}\left(\delta\boldsymbol{F}^T\left(\ddot{\boldsymbol{F}}\boldsymbol{E}_0 + 2\boldsymbol{F}DU(\boldsymbol{C}) - \boldsymbol{M}_{\text{ext}}\right)\right) = 0 \tag{131}$$

Applying the polar decomposition theorem to the deformation gradient, we get

$$\boldsymbol{F} = \boldsymbol{R}\boldsymbol{U} \qquad \text{and} \qquad \delta\boldsymbol{F} = \delta\boldsymbol{R}\boldsymbol{U} + \boldsymbol{R}\delta\boldsymbol{U} \tag{132}$$

Since $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}$, $\delta\boldsymbol{R}\boldsymbol{R}^T + \boldsymbol{R}\delta\boldsymbol{R}^T = \boldsymbol{0}$, and consequently

$$\widehat{\boldsymbol{\omega}}_\delta = \delta\boldsymbol{R}\boldsymbol{R}^T \tag{133}$$

is skew-symmetric. A straightforward calculation shows that (131) can be rewritten as

$$\operatorname{tr}\left(\delta\boldsymbol{U}\boldsymbol{U}\left(\boldsymbol{F}^{-1}\ddot{\boldsymbol{F}}\boldsymbol{E}_0 + 2DU(\boldsymbol{C}) - \boldsymbol{F}^{-1}\boldsymbol{M}_{\text{ext}}\right)\right) = 0 \tag{134}$$

$$\boldsymbol{\omega}_\delta \cdot \left(2\text{vect}\left(\ddot{\boldsymbol{F}}\boldsymbol{E}_0\boldsymbol{F}^T\right) - 2\text{vect}\left(\boldsymbol{M}_{\text{ext}}\boldsymbol{F}^T\right)\right) = 0 \tag{135}$$

Accordingly, the nine independent equations emanating from (131) have been converted to six independent equations (134) plus three independent equations (135). In (135), vect(•) denotes the vector invariant of a second-order tensor defined by

$$\text{vect}(\boldsymbol{a} \otimes \boldsymbol{b}) \times \boldsymbol{c} = \text{skew}(\boldsymbol{a} \otimes \boldsymbol{b})\boldsymbol{c}$$

Since

$$\begin{aligned} \text{skew}(\boldsymbol{a} \otimes \boldsymbol{b})\boldsymbol{c} &= \frac{1}{2}(\boldsymbol{a} \otimes \boldsymbol{b} - \boldsymbol{b} \otimes \boldsymbol{a})\boldsymbol{c} \\ &= \frac{1}{2}\left((\boldsymbol{b} \cdot \boldsymbol{c})\boldsymbol{a} - (\boldsymbol{a} \cdot \boldsymbol{c})\boldsymbol{b}\right) \\ &= \frac{1}{2}(\boldsymbol{b} \times \boldsymbol{a}) \times \boldsymbol{c} \end{aligned}$$

we have

$$\text{vect}(\boldsymbol{a} \otimes \boldsymbol{b}) = \frac{1}{2}(\boldsymbol{b} \times \boldsymbol{a}) \tag{136}$$

Accordingly,

$$2\text{vect}\big(\ddot{\boldsymbol{F}}\boldsymbol{E}_0\boldsymbol{F}^T\big) = 2\text{vect}\big(E_0^{ij}\ddot{\boldsymbol{d}}_i \otimes \boldsymbol{d}_j\big) = E_0^{ij}\boldsymbol{d}_j \times \ddot{\boldsymbol{d}}_i \tag{137}$$

$$2\text{vect}\big(\boldsymbol{M}_{\text{ext}}\boldsymbol{F}^T\big) = 2\text{vect}\big(\boldsymbol{f}^i_{\text{ext}} \otimes \boldsymbol{d}_i\big) = \boldsymbol{d}_i \times \boldsymbol{f}^i_{\text{ext}} = \overline{\boldsymbol{m}}_{\text{ext}} \tag{138}$$

A.1.1 Balance of Angular Momentum

To get the balance law for angular momentum, substitute $\delta\boldsymbol{U} = \boldsymbol{0}$ into (134), $\boldsymbol{\omega}_\delta = \boldsymbol{\xi}$ into (135), and $\delta\overline{\boldsymbol{x}} = \boldsymbol{\xi} \times \overline{\boldsymbol{x}}$ into (130). Subsequent summation of the resulting equations yields

$$\boldsymbol{\xi} \cdot \big(M\overline{\boldsymbol{x}} \times \ddot{\overline{\boldsymbol{x}}} + 2\text{vect}\big(\ddot{\boldsymbol{F}}\boldsymbol{E}_0\boldsymbol{F}^T\big) - \overline{\boldsymbol{x}} \times \boldsymbol{f}_{\text{ext}} - 2\text{vect}\big(\boldsymbol{M}_{\text{ext}}\boldsymbol{F}^T\big)\big) = 0$$

or

$$\boldsymbol{\xi} \cdot \left(\frac{d}{dt}\boldsymbol{j} - \boldsymbol{m}_{\text{ext}}\right) = 0$$

The last equation has to hold for arbitrary $\boldsymbol{\xi} \in \mathbb{R}^3$. Accordingly, one obtains $d\boldsymbol{j}/dt = \boldsymbol{m}_{\text{ext}}$, where

$$\begin{aligned} \boldsymbol{j} &= M\overline{\boldsymbol{x}} \times \dot{\overline{\boldsymbol{x}}} + 2\text{vect}\big(\dot{\boldsymbol{F}}\boldsymbol{E}_0\boldsymbol{F}^T\big) \\ \boldsymbol{m}_{\text{ext}} &= \overline{\boldsymbol{x}} \times \boldsymbol{f}_{\text{ext}} + 2\text{vect}\big(\boldsymbol{M}_{\text{ext}}\boldsymbol{F}^T\big) \end{aligned}$$

denote, respectively, the total angular momentum and the resultant external torque with respect to the origin of the inertial frame of reference. Note that the same conclusions can be drawn by substituting $\delta\bar{\boldsymbol{x}} = \boldsymbol{\xi} \times \bar{\boldsymbol{x}}$ into (130), and $\delta\boldsymbol{F} = \widehat{\boldsymbol{\xi}}\boldsymbol{F}$ into (131).

A.1.2 Balance of Energy

Suppose that an external force $\boldsymbol{f}_{\text{ext}} \in \mathbb{R}^3$ along with external director forces $\boldsymbol{f}^i_{\text{ext}} \in \mathbb{R}^3$, $i = 1, 2, 3$, are acting on the body under consideration. Recall that the external director forces $\boldsymbol{f}^i_{\text{ext}}$ can be linked to the second-order tensor $\boldsymbol{M}_{\text{ext}}$ via $\boldsymbol{M}_{\text{ext}} = \boldsymbol{f}^i_{\text{ext}} \otimes \boldsymbol{D}_i$ (see Eq. (38) in Sect. 2.2). To define the external director forces we introduce 9 independent quantities $\mathcal{M}^{ij}$ such that

$$\boldsymbol{\mathcal{M}} = \mathcal{M}^{ij}\boldsymbol{D}_i \otimes \boldsymbol{D}_j \tag{139}$$

and

$$\boldsymbol{M}_{\text{ext}} = \boldsymbol{F}\boldsymbol{\mathcal{M}} = \mathcal{M}^{ij}\boldsymbol{d}_i \otimes \boldsymbol{D}_j \tag{140}$$

Note that the last equation implies

$$\boldsymbol{f}^i_{\text{ext}} = \mathcal{M}^{ki}\boldsymbol{d}_k \tag{141}$$

Now substitute $\dot{\bar{\boldsymbol{x}}}$ for $\delta\bar{\boldsymbol{x}}$ into (130) and $\dot{\boldsymbol{F}}$ for $\delta\boldsymbol{F}$ into (131). Subsequent summation of both equations yields

$$\dot{\bar{\boldsymbol{x}}} \cdot \left(M\ddot{\bar{\boldsymbol{x}}} - \boldsymbol{f}_{\text{ext}}\right) + \operatorname{tr}\left(\dot{\boldsymbol{F}}^T\left(\ddot{\boldsymbol{F}}\boldsymbol{E}_0 + 2\boldsymbol{F}DU(\boldsymbol{C}) - \boldsymbol{M}_{\text{ext}}\right)\right) = 0 \tag{142}$$

Taking into account the relationships

$$\dot{\bar{\boldsymbol{x}}} \cdot M\ddot{\bar{\boldsymbol{x}}} = \frac{d}{dt}\left(\frac{1}{2}M\dot{\bar{\boldsymbol{x}}} \cdot \dot{\bar{\boldsymbol{x}}}\right)$$
$$\operatorname{tr}\left(\dot{\boldsymbol{F}}^T\ddot{\boldsymbol{F}}\boldsymbol{E}_0\right) = \frac{d}{dt}\left(\frac{1}{2}\operatorname{tr}\left(\dot{\boldsymbol{F}}^T\dot{\boldsymbol{F}}\boldsymbol{E}_0\right)\right)$$

we define the kinetic energy

$$T = \frac{1}{2}M\dot{\bar{\boldsymbol{x}}} \cdot \dot{\bar{\boldsymbol{x}}} + \frac{1}{2}\operatorname{tr}\left(\dot{\boldsymbol{F}}\boldsymbol{E}_0\dot{\boldsymbol{F}}^T\right) \tag{143}$$

Moreover,

$$\begin{aligned}\operatorname{tr}\left(\dot{\boldsymbol{F}}^T\boldsymbol{F}2DU(\boldsymbol{C})\right) &= \operatorname{tr}\left(2DU(\boldsymbol{C})\operatorname{sym}(\dot{\boldsymbol{F}}^T\boldsymbol{F})\right)\\ &= \operatorname{tr}\left(2DU(\boldsymbol{C})\frac{1}{2}(\dot{\boldsymbol{F}}^T\boldsymbol{F}+\boldsymbol{F}^T\dot{\boldsymbol{F}})\right)\\ &= \operatorname{tr}\left(2DU(\boldsymbol{C})\frac{1}{2}\dot{\boldsymbol{C}}\right)\\ &= \frac{d}{dt}U(\boldsymbol{C})\end{aligned}$$

Now (142) can be recast in the form

$$\frac{d}{dt}E = P_{\text{ext}} \tag{144}$$

Here, E is the total mechanical energy given by

$$E = T + U$$

where U denotes the total strain energy defined in (13). On the right hand side of balance equation (144)

$$P_{\text{ext}} = \boldsymbol{f}_{\text{ext}} \cdot \dot{\boldsymbol{x}} + \operatorname{tr}\left(\dot{\boldsymbol{F}}^T\boldsymbol{M}_{\text{ext}}\right)$$

denotes the power of the external forces acting on the pseudo-rigid body. We next focus on the power of the director forces given by

$$\overline{P}_{\text{ext}} = \operatorname{tr}\left(\dot{\boldsymbol{F}}^T\boldsymbol{M}_{\text{ext}}\right)$$

Taking into account (140), the last equation can be rewritten as

$$\begin{aligned}\overline{P}_{\text{ext}} &= \operatorname{tr}\left(\dot{\boldsymbol{F}}^T\boldsymbol{F}\boldsymbol{\mathcal{M}}\right)\\ &= \operatorname{tr}\left((\overline{\boldsymbol{\mathcal{M}}}+\widetilde{\boldsymbol{\mathcal{M}}})\dot{\boldsymbol{F}}^T\boldsymbol{F}\right)\end{aligned}$$

In the last equation

$$\begin{aligned}\overline{\boldsymbol{\mathcal{M}}} &= \operatorname{sym}(\boldsymbol{\mathcal{M}})\\ \widetilde{\boldsymbol{\mathcal{M}}} &= \operatorname{skew}(\boldsymbol{\mathcal{M}})\end{aligned}$$

have been introduced. Now

$$\operatorname{tr}\left(\overline{\boldsymbol{\mathcal{M}}}\dot{\boldsymbol{F}}^T\boldsymbol{F}\right) = \operatorname{tr}\left(\overline{\boldsymbol{\mathcal{M}}}\operatorname{sym}\left(\dot{\boldsymbol{F}}^T\boldsymbol{F}\right)\right) = \frac{1}{2}\operatorname{tr}\left(\overline{\boldsymbol{\mathcal{M}}}\dot{\boldsymbol{C}}\right)$$

Furthermore,

$$\operatorname{tr}\left(\widetilde{\boldsymbol{\mathcal{M}}}\dot{\boldsymbol{F}}^T\boldsymbol{F}\right) = \operatorname{tr}\left(\widetilde{\boldsymbol{\mathcal{M}}}\operatorname{skew}\left(\dot{\boldsymbol{F}}^T\boldsymbol{F}\right)\right)$$

Applying the polar decomposition $\boldsymbol{F} = \boldsymbol{R}\boldsymbol{U}$ along with $\dot{\boldsymbol{F}} = \dot{\boldsymbol{R}}\boldsymbol{U} + \boldsymbol{R}\dot{\boldsymbol{U}}$ and $\widehat{\boldsymbol{\omega}} = \dot{\boldsymbol{R}}\boldsymbol{R}^T$ (cf. (132) and (133) on p. 47), we get

$$\operatorname{tr}\left(\widetilde{\boldsymbol{\mathcal{M}}}\operatorname{skew}\left(\dot{\boldsymbol{F}}^T\boldsymbol{F}\right)\right) = \boldsymbol{\omega}\cdot 2\operatorname{vect}\left(\boldsymbol{F}\widetilde{\boldsymbol{\mathcal{M}}}\boldsymbol{F}^T\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{\mathcal{M}}}\dot{\boldsymbol{U}}^T\boldsymbol{U}\right)$$

Altogether the power of the external forces can be written in the form

$$\begin{aligned} P_{\text{ext}} &= \boldsymbol{f}_{\text{ext}}\cdot\dot{\bar{\boldsymbol{x}}} + \frac{1}{2}\operatorname{tr}\left(\overline{\boldsymbol{\mathcal{M}}}\dot{\boldsymbol{C}}\right) + \boldsymbol{\omega}\cdot 2\operatorname{vect}\left(\boldsymbol{F}\widetilde{\boldsymbol{\mathcal{M}}}\boldsymbol{F}^T\right) + \operatorname{tr}\left(\widetilde{\boldsymbol{\mathcal{M}}}\dot{\boldsymbol{U}}^T\boldsymbol{U}\right) \\ &= \boldsymbol{f}_{\text{ext}}\cdot\dot{\bar{\boldsymbol{x}}} + \frac{1}{2}\overline{\mathcal{M}}^{ij}\dot{d}_{ij} + \boldsymbol{\omega}\cdot\overline{\boldsymbol{m}}_{\text{ext}} + \operatorname{tr}\left(\widetilde{\boldsymbol{\mathcal{M}}}\dot{\boldsymbol{U}}^T\boldsymbol{U}\right) \end{aligned}$$

Here, $\overline{\boldsymbol{m}}_{\text{ext}}$ can be identified as the resultant external torque relative to the center of mass that has been introduced in (71). In particular, we have

$$\begin{aligned} \overline{\boldsymbol{m}}_{\text{ext}} &= 2\operatorname{vect}\left(\boldsymbol{F}\widetilde{\boldsymbol{\mathcal{M}}}\boldsymbol{F}^T\right) \\ &= \widetilde{\mathcal{M}}^{ji}\boldsymbol{d}_i\times\boldsymbol{d}_j \\ &= \varepsilon_{ijk}\widetilde{\mathcal{M}}^{ji}\boldsymbol{d}^k \\ &= \boldsymbol{d}_i\times\boldsymbol{f}^i_{\text{ext}} \end{aligned}$$

In the last equation use has been made of (141). Moreover,

$$\varepsilon_{ijk} = (\boldsymbol{d}_i\times\boldsymbol{d}_j)\cdot\boldsymbol{d}_k = e_{ijk}\sqrt{d}$$

where $d = \det(d_{ij})$ (or $\sqrt{d} = (\boldsymbol{d}_1\times\boldsymbol{d}_2)\cdot\boldsymbol{d}_3$) and e_{ijk} denotes the alternating symbol.

A.2 Application of External Torques

It can be observed from the above treatment that the application of external torques $\overline{\boldsymbol{m}}_{\text{ext}}$ relative to the center of mass is linked to the skew-symmetric tensor $\widetilde{\boldsymbol{\mathcal{M}}} = \widetilde{\mathcal{M}}^{ij}\boldsymbol{D}_i\otimes\boldsymbol{D}_j$. In particular, given the covariant components of the external torque, $m_k = \boldsymbol{d}_k\cdot\overline{\boldsymbol{m}}_{\text{ext}}$, we obtain

$$m_k = \varepsilon_{ijk}\widetilde{\mathcal{M}}^{ji}$$

from which it follows that

$$\begin{aligned}\widetilde{\mathcal{M}}^{23} &= -\widetilde{\mathcal{M}}^{32} = -\frac{m_1}{2\sqrt{d}}\\ \widetilde{\mathcal{M}}^{31} &= -\widetilde{\mathcal{M}}^{13} = -\frac{m_2}{2\sqrt{d}}\\ \widetilde{\mathcal{M}}^{12} &= -\widetilde{\mathcal{M}}^{21} = -\frac{m_3}{2\sqrt{d}}\end{aligned}$$

or

$$\widetilde{\mathcal{M}}^{ij} = \frac{e^{jik} m_k}{2\sqrt{d}} \tag{145}$$

where $e^{ijk} = e_{ijk}$ again denotes the alternating symbol. Accordingly, using the above formulas for $\widetilde{\mathcal{M}}^{ij}$ in terms of the torque components m_k, the corresponding director forces can be calculated via

$$\boldsymbol{f}^j_{\text{ext}} = \widetilde{\mathcal{M}}^{ij} \boldsymbol{d}_i \tag{146}$$

or

$$\boldsymbol{f}^j_{\text{ext}} = \frac{e^{ijk}}{2\sqrt{d}} \left(\boldsymbol{d}_j \otimes \boldsymbol{d}_k\right) \overline{\boldsymbol{m}}_{\text{ext}} \tag{147}$$

To summarize, the action of an external torque $\overline{\boldsymbol{m}}_{\text{ext}}$ relative to the center of mass can be realized by applying external director forces of the form

$$\begin{bmatrix}\boldsymbol{f}^1_{\text{ext}}\\ \boldsymbol{f}^2_{\text{ext}}\\ \boldsymbol{f}^3_{\text{ext}}\end{bmatrix} = \frac{1}{2\sqrt{d}} \begin{bmatrix}\boldsymbol{d}_2 \otimes \boldsymbol{d}_3 - \boldsymbol{d}_3 \otimes \boldsymbol{d}_2\\ \boldsymbol{d}_3 \otimes \boldsymbol{d}_1 - \boldsymbol{d}_1 \otimes \boldsymbol{d}_3\\ \boldsymbol{d}_1 \otimes \boldsymbol{d}_2 - \boldsymbol{d}_2 \otimes \boldsymbol{d}_1\end{bmatrix} \overline{\boldsymbol{m}}_{\text{ext}} \tag{148}$$

Remark A.1 Formula (148) can be viewed as an extension to flexible Cosserat points of the method proposed in Betsch and Sänger (2013). In this work the consistent application of torques has been dealt with in the context of rigid body dynamics formulated in terms of directors (or direction cosines). The formula proposed in Betsch and Sänger (2013) is given by

$$\boldsymbol{f}^j_{\text{ext}} = \frac{1}{2} \overline{\boldsymbol{m}}_{\text{ext}} \times \boldsymbol{d}^j \tag{149}$$

The equivalence of (149) to (148) can be shown by a direct calculation:

$$\begin{aligned}\boldsymbol{f}^j_{\text{ext}} &= \frac{1}{2}\overline{\boldsymbol{m}}_{\text{ext}} \times \boldsymbol{d}^j \\ &= \frac{1}{2}m_k\boldsymbol{d}^k \times \boldsymbol{d}^j \\ &= \frac{1}{2}m_k d^{-\frac{1}{2}} e^{kji}\boldsymbol{d}_i \\ &= \widetilde{\mathcal{M}}^{ij}\boldsymbol{d}_i\end{aligned}$$

where (145) has been used.

A.2.1 Fully Actuated Cosserat Point

If the Cosserat point shall be fully actuated, the 9 independent quantities $\mathcal{M}^{ij}$ in (139) can be employed as control inputs. According to (141) this approach determines the external director forces

$$\boldsymbol{f}^i_{\text{ext}} = \mathcal{M}^{ji}\boldsymbol{d}_j$$

If required the external torque associated with the control inputs can be extracted via

$$\overline{\boldsymbol{m}}_{\text{ext}} = \mathcal{M}^{ji}\boldsymbol{d}_i \times \boldsymbol{d}_j$$

Note that due to the presence of the cross product the skew-symmetric part of $\mathcal{M}^{ji}$, that is, $\widetilde{\mathcal{M}}^{ji} = (\mathcal{M}^{ji} - \mathcal{M}^{ij})/2$, is automatically extracted. The above result coincides with

$$\overline{\boldsymbol{m}}_{\text{ext}} = m_k\boldsymbol{d}^k \quad \text{where} \quad m_k = \sqrt{d}\, e_{ijk}\mathcal{M}^{ji}$$

Again the skew-symmetric part of $\mathcal{M}^{ji}$ is extracted due to the presence of the alternating symbol.

A.3 Iteration Matrix of the EM Integrator

Consider St. Venant-Kirchhoff material with strain energy density

$$W(\boldsymbol{G}) = \frac{\lambda}{2}(\text{tr}\boldsymbol{G})^2 + \mu\text{tr}\left(\boldsymbol{G}^2\right)$$

where the Green–Lagrangean strain tensor is given by

$$\boldsymbol{G} = \frac{1}{2}(\boldsymbol{C} - \boldsymbol{I}) = \gamma_{ij}\boldsymbol{D}^i \otimes \boldsymbol{D}^j$$

Note that the components $\gamma_{ij} = \frac{1}{2}(d_{ij} - \delta_{ij})$ have been introduced in (66). According to (11), the second Piola-Kirchhoff stress tensor is given by

$$\begin{aligned} \boldsymbol{S} &= 2DW(\boldsymbol{C}) \\ &= DW(\boldsymbol{G}) \\ &= \lambda\,(\mathrm{tr}\boldsymbol{G})\,\boldsymbol{I} + 2\mu\boldsymbol{G} \end{aligned} \tag{150}$$

Moreover, the fourth-order elasticity tensor assumes the form

$$\begin{aligned} \mathsf{C} &= 4D^2W(\boldsymbol{C}) \\ &= D^2W(\boldsymbol{G}) \\ &= \lambda\boldsymbol{I} \otimes \boldsymbol{I} + 2\mu\mathsf{I} \end{aligned} \tag{151}$$

Since we have assumed that the director triad $\{\boldsymbol{D}_i\}$ in the reference configuration is orthonormal, it suffices to consider the Cartesian components of $\boldsymbol{S}$ and C. Accordingly, we have

$$S_{ij} = \lambda\gamma_{kk}\delta_{ij} + 2\mu\gamma_{ij} \tag{152}$$

and

$$\mathsf{C}_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)$$

We next deal with the linearization of the internal director forces $\boldsymbol{f}^i_{\mathrm{int}} = S^{ij}\boldsymbol{d}_j$. First consider the time-continuous case where, according to the product rule of differentiation, we get

$$\Delta\boldsymbol{f}^i_{\mathrm{int}} = \Delta S^{ij}\boldsymbol{d}_j + S^{ij}\Delta\boldsymbol{d}_j \tag{153}$$

With regard to (152)

$$\begin{aligned} \Delta S_{ij} &= \lambda\Delta\gamma_{kk}\delta_{ij} + 2\mu\Delta\gamma_{ij} \\ &= \lambda(\boldsymbol{d}_k \cdot \Delta\boldsymbol{d}_k)\delta_{ij} + \mu(\boldsymbol{d}_i \cdot \Delta\boldsymbol{d}_j + \boldsymbol{d}_j \cdot \Delta\boldsymbol{d}_i) \end{aligned} \tag{154}$$

Now, a straightforward calculation yields

$$\Delta\boldsymbol{f}^i_{\mathrm{int}} = \left(\boldsymbol{K}^{ij}_{\mathrm{mat}} + \boldsymbol{K}^{ij}_{\mathrm{geo}}\right)\Delta\boldsymbol{d}_j \tag{155}$$

where the contributions to the iteration matrix have been split into a material part $\boldsymbol{K}^{ij}_{\mathrm{mat}}$ and a geometric part $\boldsymbol{K}^{ij}_{\mathrm{geo}}$. The material part is given by

$$\boldsymbol{K}^{ij}_{\mathrm{mat}} = \lambda\boldsymbol{d}_i \otimes \boldsymbol{d}_j + \mu\boldsymbol{d}_j \otimes \boldsymbol{d}_i + \mu\boldsymbol{d}_k \otimes \boldsymbol{d}_k\,\delta_{ij}$$

and the geometric part assumes the form

$$\boldsymbol{K}^{ij}_{\mathrm{geo}} = S_{ij}\boldsymbol{I}$$

The symmetry of the iteration matrix follows from the properties $\boldsymbol{K}^{ij}_{\text{mat}} = (\boldsymbol{K}^{ji}_{\text{mat}})^T$ and $\boldsymbol{K}^{ij}_{\text{geo}} = (\boldsymbol{K}^{ji}_{\text{geo}})^T$. Similar to (153), in the discrete case the EM scheme (63) leads to

$$\Delta\ (\boldsymbol{f}^i_{\text{int}})\big|_{n+\frac{1}{2}} = \Delta S^{ij}_A \boldsymbol{d}_{j_{n+\frac{1}{2}}} + S^{ij}_A \Delta \boldsymbol{d}_{j_{n+\frac{1}{2}}}$$

Similar to (154), the algorithmic stress formula (65) leads to

$$\begin{aligned} \Delta S^{ij}_A &= \lambda \Delta \gamma_{kk_{n+\frac{1}{2}}} \delta_{ij} + 2\mu \Delta \gamma_{ij_{n+\frac{1}{2}}} \\ &= \tfrac{\lambda}{2} (\boldsymbol{d}_{k_{n+1}} \cdot \Delta \boldsymbol{d}_{k_{n+1}}) \delta_{ij} + \tfrac{\mu}{2} (\boldsymbol{d}_{i_{n+1}} \cdot \Delta \boldsymbol{d}_{j_{n+1}} + \boldsymbol{d}_{j_{n+1}} \cdot \Delta \boldsymbol{d}_{i_{n+1}}) \end{aligned} \tag{156}$$

Altogether the discrete counterpart of (155) is given by the consistent linearization

$$\Delta\ (\boldsymbol{f}^i_{\text{int}})\big|_{n+\frac{1}{2}} = \left(\boldsymbol{K}^{ij}_{\text{mat}}\Big|_{n+\frac{1}{2}} + \boldsymbol{K}^{ij}_{\text{geo}}\Big|_{n+\frac{1}{2}} \right) \Delta \boldsymbol{d}_{j_{n+1}}$$

where the material part is given by

$$\boldsymbol{K}^{ij}_{\text{mat}}\Big|_{n+\frac{1}{2}} = \frac{1}{2} \left(\lambda \boldsymbol{d}_{i_{n+\frac{1}{2}}} \otimes \boldsymbol{d}_{j_{n+1}} + \mu \boldsymbol{d}_{j_{n+\frac{1}{2}}} \otimes \boldsymbol{d}_{i_{n+1}} + \mu \boldsymbol{d}_{k_{n+\frac{1}{2}}} \otimes \boldsymbol{d}_{k_{n+1}}\, \delta_{ij} \right)$$

and the geometric part assumes the form

$$\boldsymbol{K}^{ij}_{\text{geo}}\Big|_{n+\frac{1}{2}} = \frac{1}{2} S^{ij}_A \boldsymbol{I}$$

It is obvious that in the discrete setting the material part destroys the symmetry of the iteration matrix, for

$$\left(\boldsymbol{K}^{ij}_{\text{mat}} \right)\Big|_{n+\frac{1}{2}} \neq \left(\boldsymbol{K}^{ji}_{\text{mat}} \right)^T\Big|_{n+\frac{1}{2}}$$

We finally remark that due to definition (13) of the total strain energy of the Cosserat point, namely $U(\boldsymbol{C}) = V_0 W(\boldsymbol{C})$, the above stress components S^{ij} should be replaced by $\overline{S}^{ij} = V_0 S^{ij}$.

References

Antman, S. S. (2005). *Nonlinear problems of elasticity* (2nd ed.). Springer.

Armero, F., & Petöcz, E. (1998). Formulation and analysis of conserving algorithms for frictionless dynamic contact/impact problems. *Computer Methods in Applied Mechanics and Engineering*, *158*, 269–300.

Armero, F., & Romero, I. (2001). On the formulation of high-frequency dissipative time-stepping algorithms for nonlinear dynamics. Part II: Second-order methods. *Computer Methods in Applied Mechanics and Engineering*, *190*, 6783–6824.

Armero, F., & Romero, I. (2003). Energy-dissipative momentum-conserving time-stepping algorithms for the dynamics of nonlinear Cosserat rods. *Computational Mechanics*, *31*, 3–26.

Ascher, U. M., & Petzold, L. R. (1998). Computer methods for ordinary differential equations and differential-algebraic equations. SIAM.

Bauchau, O. A. (2011). *Flexible multibody dynamics*. Solid mechanics and its applications New York: Springer.

Bauchau, O. A., & Bottasso, C. L. (1999). On the design of energy preserving and decaying schemes for flexible, nonlinear multi-body systems. *Computer Methods in Applied Mechanics and Engineering*, *169*(1–2), 61–79.

Betsch, P. (2006). Energy-consistent numerical integration of mechanical systems with mixed holonomic and nonholonomic constraints. *Computer Methods in Applied Mechanics and Engineering*, *195*, 7020–7035.

Betsch, P., & Leyendecker, S. (2006). The discrete null space method for the energy consistent integration of constrained mechanical systems. Part II: Multibody dynamics. *International Journal for Numerical Methods in Engineering*, *67*(4), 499–552.

Betsch, P., & Sänger, N. (2009a). On the use of geometrically exact shells in a conserving framework for flexible multibody dynamics. *Computer Methods in Applied Mechanics and Engineering*, *198*, 1609–1630.

Betsch, P., & Sänger, N. (2009b). A nonlinear finite element framework for flexible multibody dynamics: Rotationless formulation and energy-momentum conserving discretization. In C. L. Bottasso (Ed.), *Multibody Dynamics: Computational Methods and Applications, Computational Methods in Applied Sciences* (Vol. 12, pp. 119–141). Springer.

Betsch, P., & Sänger, N. (2013). On the consistent formulation of torques in a rotationless framework for multibody dynamics. *Computers & Structures*, *127*, 29–38.

Betsch, P., & Steinmann, P. (2002a). Conservation properties of a time FE method. Part III: Mechanical systems with holonomic constraints. *International Journal for Numerical Methods in Engineering*, *53*, 2271–2304.

Betsch, P., & Steinmann, P. (2002b). Frame-indifferent beam finite elements based upon the geometrically exact beam theory. *International Journal for Numerical Methods in Engineering*, *54*, 1775–1788.

Betsch, P., & Steinmann, P. (2002c). A DAE approach to flexible multibody dynamics. *Multibody System Dynamics*, *8*, 367–391.

Betsch, P., & Steinmann, P. (2003). Constrained dynamics of geometrically exact beams. *Computational Mechanics*, *31*, 49–59.

Betsch, P., & Uhlar, S. (2007). Energy-momentum conserving integration of multibody dynamics. *Multibody System Dynamics*, *17*(4), 243–289.

Betsch, P., Hesch, C., Sänger, N., & Uhlar, S. (2010). Variational integrators and energy-momentum schemes for flexible multibody dynamics. *Journal of Computational and Nonlinear Dynamics*, *5*(3), 031001/1-11.

Betsch, P., Siebert, R., & Sänger, N. (2012). Natural coordinates in the optimal control of multibody systems. *Journal of Computational and Nonlinear Dynamics*, *7*(1), 011009/1-8.

Bottasso, C. L., & Croce, A. (2004). Optimal control of multibody systems using an energy preserving direct transcription method. *Multibody System Dynamics*, *12*(1), 17–45.

Bottasso, C. L., & Trainelli, L. (2004). An attempt at the classification of energy decaying schemes for structural and multibody dynamics. *Multibody System Dynamics*, *12*(2), 173–185.

Bottasso, C. L., Borri, M., & Trainelli, L. (2001). Integration of elastic multibody systems by invariant conserving/dissipating algorithms. II. Numerical schemes and applications. *Computer Methods in Applied Mechanics and Engineering*, *190*, 3701–3733.

Bottasso, C. L., Bauchau, O. A., & Choi, J.-Y. (2002). An energy decaying scheme for nonlinear dynamics of shells. *Computer Methods in Applied Mechanics and Engineering*, *191*(27–28), 3099–3121.

Bottema, O., & Roth, B. (1979). *Theoretical Kinematics*. Amsterdam: North-Holland Publishing Company.

Brank, B., Briseghella, L., Tonello, N., & Damjanic, F. B. (1998). On non-linear dynamics of shells: Implementation of energy-momentum conserving algorithm for a finite rotation shell model. *International Journal for Numerical Methods in Engineering*, *42*, 409–442.

Cohen, H., & Muncaster, R. G. (1988). *The theory of Pseudo-rigid Bodies*. New York: Springer.

Conde Martín, S., Betsch, P., & García Orden, J. C. (2016). A temperature-based thermodynamically consistent integration scheme for discrete thermo-elastodynamics. *Communications in Nonlinear Science and Numerical Simulation*, *32*, 63–80.

Crisfield, M. A. (1997). *Non-linear finite element analysis of solids and structures*. Advanced topics. New York: Wiley.

Crisfield, M. A., & Shi, J. (1994). A co-rotational element/time-integration strategy for non-linear dynamics. *International Journal for Numerical Methods in Engineering*, *37*, 1897–1913.

de García Jalón, J. (2007). Twenty-five years of natural coordinates. *Multibody System Dynamics*, *18*(1), 15–33.

de Jalón, J. G., & Bayo, E. (1994). *Kinematic and dynamic simulation of multibody systems: The real-time challenge*. Springer.

Géradin, M. G., & Cardona, A. (2001). *Flexible multibody dynamics: A finite element approach*. Wiley.

Gonzalez, O. (1996). Time integration and discrete Hamiltonian systems. *Journal of Nonlinear Science*, *6*, 449–467.

Gonzalez, O., & Simo, J. C. (1996). On the stability of symplectic and energy-momentum algorithms for non-linear Hamiltonian systems with symmetry. *Computer Methods in Applied Mechanics and Engineering*, *134*, 197–222.

Greenspan, D. (1984). Conservative numerical methods for $\ddot{x} = f(x)$. *Journal of Computational Physics*, *56*, 28–41.

Groß, M., & Betsch, P. (2011). Galerkin-based energy-momentum consistent time-stepping algorithms for classical nonlinear thermo-elastodynamics. *Mathematics and Computers in Simulation*, *82*(4), 718–770.

Groß, M., & Betsch, P. (2010). Energy-momentum consistent finite element discretization of dynamic finite viscoelasticity. *International Journal for Numerical Methods in Engineering*, *81*(11), 1341–1386.

Groß, M., Betsch, P., & Steinmann, P. (2005). Conservation properties of a time FE method. Part IV: Higher order energy and momentum conserving. *International Journal for Numerical Methods in Engineering*, *63*, 1849–1897.

Gurtin, M. E. (1981). *An introduction to continuum mechanics*. Academic Press.

Hesch, C., & Betsch, P. (2011a). Transient 3d contact problems-NTS method: mixed methods and conserving integration. *Computational Mechanics*, *48*(4), 437–449.

Hesch, C., & Betsch, P. (2010). Transient three-dimensional domain decomposition problems: Frame-indifferent mortar constraints and conserving integration. *International Journal for Numerical Methods in Engineering*, *82*(3), 329–358.

Hesch, C., & Betsch, P. (2011b). Transient three-dimensional contact problems: mortar method. Mixed methods and conserving integration. *Computational Mechanics*, *48*(4), 461–475.

Hesch, C., & Betsch, P. (2009). A mortar method for energy-momentum conserving schemes in frictionless dynamic contact problems. *International Journal for Numerical Methods in Engineering*, *77*(10), 1468–1500.

Hesch, C., & Betsch, P. (2011c). Energy-momentum consistent algorithms for dynamic thermomechanical problems—application to mortar domain decomposition problems. *International Journal for Numerical Methods in Engineering*, *86*(11), 1277–1302.

Hughes, T. J. R. (2000). *The Finite element method*. Dover Publications.

Hughes, T. J. R., & Winget, J. (1980). Finite rotation effects in numerical integration of rate constitutive equations arising in large-deformation analysis. *International Journal for Numerical Methods in Engineering*, *15*, 1862–1867.

Hughes, T. J. R., Caughey, T. K., & Liu, W. K. (1978). Finite-element methods for nonlinear elastodynamics which conserve energy. *Journal of Applied Mechanics*, *45*, 366–370.

Ibrahimbegović, A. (2009). *Nonlinear solid mechanics. Solid mechanics and its applications* (Vol. 160). Springer.

Ibrahimbegović, A., Mamouri, S., Taylor, R. L., & Chen, A. J. (2000). Finite element method in dynamics of flexible multibody systems: Modeling of holonomic constraints and energy conserving integration schemes. *Multibody System Dynamics*, *4*(2–3), 195–223.

Johnson, E. R., & Murphey, T. D. (2009). Scalable variational integrators for constrained mechanical systems in generalized coordinates. *IEEE Transactions on Robotics*, *25*(6), 1249–1261.

Koch, M. W., & Leyendecker, S. (2013). Energy momentum consistent force formulation for the optimal control of multibody systems. *Multibody System Dynamics*, *29*, 381–401.

Krenk, S. (2009). *Non-linear modeling and analysis of solids and structures*. Cambridge University Press.

Kuhl, D., & Crisfield, M. A. (1999). Energy-conserving and decaying algorithms in non-linear structural mechanics. *International Journal for Numerical Methods in Engineering*, *45*, 569–599.

Kunkel, P., & Mehrmann, V. (2006). *Differential-algebraic equations*. European Mathematical Society.

Laursen, T. A. (2002). *Computational contact and impact mechanics*. Springer.

Laursen, T. A., & Chawla, V. (1997). Design of energy conserving algorithms for frictionless dynamic contact problems. *International Journal for Numerical Methods in Engineering*, *40*, 863–886.

Lens, E., & Cardona, A. (2007). An energy preserving/decaying scheme for nonlinearly constrained multibody systems. *Multibody System Dynamics*, *18*(3), 435–470.

Lens, E. V., Cardona, A., & Géradin, M. (2004). Energy preserving time integration for constrained multibody systems. *Multibody System Dynamics*, *11*(1), 41–61.

Lewis, D., & Simo, J. C. (1994). Conserving algorithms for the dynamics of Hamiltonian systems on Lie groups. *Journal of Nonlinear Science*, *4*, 253–299.

Leyendecker, S., Betsch, P., & Steinmann, P. (2006). Objective energy-momentum conserving integration for the constrained dynamics of geometrically exact beams. *Computer Methods in Applied Mechanics and Engineering*, *195*, 2313–2333.

Leyendecker, S., Betsch, P., & Steinmann, P. (2008a). The discrete null space method for the energy consistent integration of constrained mechanical systems. Part III: Flexible multibody dynamics. *Multibody System Dynamics*, *19*(1–2), 45–72.

Leyendecker, S., Marsden, J. E., & Ortiz, M. (2008b). Variational integrators for constrained dynamical systems. *Zeitschrift für Angewandte Mathematik und Mechanik (ZAMM)*, *88*(9), 677–708.

Leyendecker, S., Ober-Blöbaum, S., Marsden, J. E., & Ortiz, M. (2010). Discrete mechanics and optimal control for constrained systems. *Optimal Control Applications and Methods*, *31*(6), 505–528.

Marsden, J. E., & Ratiu, T. S. (1999). *Introduction to mechanics and symmetry* (2nd ed.). Springer.

McPhee, J. J., & Redmond, S. M. (2006). Modelling multibody systems with indirect coordinates. *Computer Methods in Applied Mechanics and Engineering*, *195*, 6942–6957.

Nordenholz, T. R., & O'Reilly, O. M. (1998). On steady motions of isotropic, elastic Cosserat points. *IMA Journal of Applied Mathematics*, *60*, 55–72.

Ober-Blöbaum, S., Junge, O., & Marsden, J. E. (2011). Discrete mechanics and optimal control: An analysis. *ESAIM: Control, Optimisation and Calculus of Variations*, *17*(2), 322–352.

Romero, I. (2009). Thermodynamically consistent time-stepping algorithms for non-linear thermomechanical systems. *International Journal for Numerical Methods in Engineering*, *79*(6), 706–732.

Romero, I. (2010). Algorithms for coupled problems that preserve symmetries and the laws of thermodynamics: Part II: Fractional step methods. *Computer Methods in Applied Mechanics and Engineering*, *199*(33–36), 2235–2248.

Romero, I. (2012). An analysis of the stress formula for energy-momentum methods in nonlinear elastodynamics. *Computational Mechanics*, *50*, 603–610.

Romero, I., & Armero, F. (2002a). An objective finite element approximation of the kinematics of geometrically exact rods and its use in the formulation of an energy-momentum conserving scheme in dynamics. *International Journal for Numerical Methods in Engineering*, *54*, 1683–1716.

Romero, I., & Armero, F. (2002b). Numerical integration of the stiff dynamics of geometrically exact shells: an energy-dissipative momentum-conserving scheme. *International Journal for Numerical Methods in Engineering*, *54*, 1043–1086.

Rubin, M. B. (2000). *Cosserat theories: shells, rods and points, solid mechanics and its applications* (Vol. 79). Kluwer Academic Publishers.

Simo, J. C., & Tarnow, N. (1992). The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics. *Zeitschrift für angewandte Mathematik und Physik (ZAMP)*, *43*, 757–792.

Simo, J. C., & Tarnow, N. (1994). A new energy and momentum conserving algorithm for the nonlinear dynamics of shells. *International Journal for Numerical Methods in Engineering*, *37*, 2527–2549.

Simo, J. C., & Wong, K. K. (1991). Unconditionally stable algorithms for rigid body dynamics that exactly preserve energy and momentum. *International Journal for Numerical Methods in Engineering*, *31*, 19–52.

Simo, J. C., Lewis, D., & Marsden, J. E. (1991). Stability of relative equilibria. Part I: The reduced energy-momentum method. *Archive for Rational Mechanics and Analysis*, *115*, 15–59.

Simo, J. C., Rifai, M. S., & Fox, D. D. (1992a). On a stress resultant geometrically exact shell model. Part VI: Conserving algorithms for non-linear dynamics. *International Journal for Numerical Methods in Engineering*, *34*, 117–164.

Simo, J. C., Tarnow, N., & Wong, K. K. (1992b). Exact energy-momentum conserving algorithms and symplectic schemes for nonlinear dynamics. *Computer Methods in Applied Mechanics and Engineering*, *100*, 63–116.

Tarnow, N. (1993). Energy and Momentum Conserving Algorithms for Hamiltonian Systems in the Nonlinear Dynamics of Solids. Ph.D. Dissertation, Sudam report no. 93–4. Stanford University.

Truesdell, C., & Noll, W. (2004). *The non-linear field theories of mechanics* (3rd ed.). Springer (2004).

A Lie Algebra Approach to Lie Group Time Integration of Constrained Systems

Martin Arnold, Alberto Cardona and Olivier Brüls

Abstract Lie group integrators preserve by construction the Lie group structure of a nonlinear configuration space. In multibody dynamics, they support a representation of (large) rotations in a Lie group setting that is free of singularities. The resulting equations of motion are differential equations on a manifold with tangent spaces being parametrized by the corresponding Lie algebra. In the present paper, we discuss the time discretization of these equations of motion by a generalized-α Lie group integrator for constrained systems and show how to exploit in this context the linear structure of the Lie algebra. This linear structure allows a very natural definition of the generalized-α Lie group integrator, an efficient practical implementation and a very detailed error analysis. Furthermore, the Lie algebra approach may be combined with analytical transformations that help to avoid an undesired order reduction phenomenon in generalized-α time integration. After a tutorial-like step-by-step introduction to the generalized-α Lie group integrator, we investigate its convergence behaviour and develop a novel initialization scheme to achieve second-order accuracy in the application to constrained systems. The theoretical results are illustrated by a comprehensive set of numerical tests for two Lie group formulations of a rotating heavy top.

1 Introduction

Structure-preserving integrators overcome limitations of classical time integration methods from the fields of ordinary differential equations (ODEs) and differential-algebraic equations (DAEs). They are known for their favourable nonlinear stability

M. Arnold (✉)
Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
e-mail: martin.arnold@mathematik.uni-halle.de

A. Cardona
Universidad Nacional Litoral - CONICET, Santa Fe, Argentina

O. Brüls
University of Liège, Liège, Belgium

P. Betsch (ed.), *Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics*, CISM International Centre for Mechanical Sciences 565, DOI 10.1007/978-3-319-31879-0_3

properties for the long-term integration of conservative systems, see, e.g., (Hairer et al. 2006).

The focus of the present paper is slightly different since we consider a class of time integration methods that is tailored to flexible multibody system models with dissipative terms resulting, e.g., from friction forces or control structures. The methods are applied to constrained systems with a nonlinear configuration space with Lie group structure. They preserve this structural property of the equations of motion in the sense that the numerical solution remains by construction in this nonlinear configuration space.

The Lie group setting allows a representation of (large) rotations that is globally free of singularities. Local parametrizations could be used to transform the system in each time step in a linear configuration space such that classical time integration methods could be used. As an alternative to such local parametrizations, Simo and Vu-Quoc (1988) proposed a Newmark-type method that is directly based on the equations of motion in a nonlinear configuration space with Lie group structure.

Starting with the work of Crouch and Grossman (1993) and Munthe-Kaas (1995, 1998), the time discretization of ordinary differential equations on Lie groups has found much interest in the numerical analysis community. This work was summarized in the comprehensive survey paper by Iserles et al. (2000). In that time, the application of Lie group time integration methods to multibody system models was studied, e.g., by Bottasso and Borri (1998) and Celledoni and Owren (2003).

In 2010, the combination of Lie group time integration with the time discretization by generalized-α methods was proposed, see (Brüls and Cardona 2010). Generalized-α methods are Newmark type methods that go back to the work of Chung and Hulbert (1993). They may be considered as a generalization of Hilber–Hughes–Taylor (HHT) methods, see (Hilber et al. 1977), and have found new interest in industrial multibody system simulation since they avoid the very strong damping of high-frequency solution components that is characteristic of other integrators in this field, see, e.g., (Negrut et al. 2005; Lunk and Simeon 2006; Jay and Negrut 2007, 2008; Arnold and Brüls 2007).

Cardona and Géradin (1994) investigated systematically the stability and convergence of HHT methods for constrained systems. This analysis may be extended to generalized-α methods, see Géradin and Cardona (2001, Sect. 10.5), and shows a risk of order reduction and large transient errors in the Lagrange multipliers and constrained forces. Numerical test results for the generalized-α Lie group integrator illustrate that this undesired numerical effect is strongly related to the specific Lie group formulation of the equations of motion, see (Brüls et al. 2011).

Therefore, the error analysis for the Lie group integrator has to consider the global errors in long-term integration as well as the transient behaviour of the numerical solution. In a series of papers, we developed a strategy for defining, implementing and analysing the Lie group integrator that is based on the observation that the increments of the configuration variables in each time step are parametrized by elements of the Lie algebra, i.e., by elements of a *linear* space, see (Arnold et al. 2011b, 2014, 2015) and (Brüls et al. 2011, 2012). In the present paper, we follow this *Lie algebra approach* and consider local and global discretization errors of the Lie group integrator as elements of the corresponding Lie algebra.

We introduce the Lie group setting in a tutorial like style and show how to discretize the equations of motion by a generalized-α Lie group integrator. There is a specific focus on practical aspects like corrector iteration and initialization of the integrator. In a comprehensive numerical test series, we consider different Lie group formulations of a heavy top benchmark problem. For the convergence analysis, we follow to a large extent the presentation in the recently published paper (Arnold et al. 2015).

The remaining part of the paper is organized as follows: Basic aspects of Lie group theory in the context of multibody dynamics and the equations of motion of constrained systems are introduced in Sect. 2. Furthermore, we discuss two different Lie group formulations of a rotating heavy top that will be used as benchmark problem throughout the paper.

In Sect. 3, we consider the generalized-α Lie group DAE integrator and study its asymptotic behaviour for time step sizes $h \to 0$. Classical results of Hilber and Hughes (1978) on "overshooting" of Newmark type methods in the application to linear problems with high-frequency solutions are shown to result in an order reduction phenomenon for the constrained case, see (Cardona and Géradin 1994). In Sect. 3.3, the large first-order error terms are illustrated by numerical tests for the heavy top benchmark problem. They may be reduced drastically by index reduction and a modification of the generalized-α Lie group integrator that is based on the so-called stabilized index-2 formulation of the equations of motion, see Sect. 3.4. Implementation aspects and some technical details are discussed in Sects. 3.5 and 3.6.

For the convergence analysis, we discuss in Sect. 4.1 a one-step error recursion of generalized-α methods for constrained systems. The coupled error propagation in differential and algebraic solution components may be studied extending the convergence analysis of ODE one-step methods to the Lie group DAE case, see Sect. 4.2. The convergence theorem for the generalized-α Lie group DAE integrators is given in Sect. 4.3. It provides the basis for an optimal initialization using perturbed starting values that guarantee second-order convergence in all solution components such that order reduction may be avoided.

2 Constrained Systems in a Configuration Space with Lie Group Structure

The main interest of this paper is in time integration methods for constrained mechanical systems that have a configuration space with Lie group structure. In the present section, we introduce this Lie group setting by studying the configuration space of a rigid body (Sect. 2.1). Lie groups are differentiable manifolds that are in a very natural way parametrized locally by elements of the corresponding Lie algebra (Sect. 2.2).

Lie groups may be used to represent large rotations in $\mathbb{R}^3$ without singularities. They are part of the mathematical framework for a generic finite element approach to flexible multibody dynamics that has been applied successfully for more than two decades (Géradin and Cardona 1989, 2001). In Sect. 2.3, we consider constrained systems and discuss the general structure of the equations of motion. As a typical example, two different Lie group formulations of a heavy top benchmark problem are introduced in Sect. 2.4. Finally, some technical details of the Lie group setting are discussed in Sect. 2.5.

2.1 The Configuration Space of a Rigid Body in $\mathbb{R}^3$

The position of a rigid body in an inertial frame is represented by a vector $\mathbf{x} \in \mathbb{R}^3$, i.e., by an element of a linear space. There are three additional degrees of freedom that describe the orientation of this rigid body but these degrees of freedom may not be represented globally by elements of a three-dimensional linear space. In engineering, small deviations from a nominal state are often characterized by three angles of rotation like Euler angles or Bryant angles (Géradin and Cardona 2001, Sect. 4.8) that suffer, however, from singularities in the case of large rotations.

Alternative representations that are free of singularities are provided, e.g., by Euler parameters that are also known as quaternions (Betsch and Siebert 2009 and Géradin and Cardona 2001, Sect. 4.5) or by the rotation matrix

$$\mathbf{R} \in \mathrm{SO}(3) := \{\, \mathbf{R} \in \mathbb{R}^{3\times 3} \;:\; \mathbf{R}^\top \mathbf{R} = \mathbf{I}_3\,,\;\; \det \mathbf{R} = +1 \,\}\,.$$

The set SO(3) is a three-dimensional differentiable manifold in $\mathbb{R}^{3\times 3}$ and may be combined in two alternative ways with the linear space $\mathbb{R}^3$ to describe the configuration of the rigid body by an element $q := (\mathbf{R}, \mathbf{x})$ of a six-dimensional group G (Brüls et al. 2011; Müller and Terze 2014a): In the direct product $G = \mathrm{SO}(3) \times \mathbb{R}^3$, the group operation $\circ$ is defined by

$$(\mathbf{R}_a, \mathbf{x}_a) \circ (\mathbf{R}_b, \mathbf{x}_b) = (\mathbf{R}_a \mathbf{R}_b, \mathbf{x}_a + \mathbf{x}_b)$$

and results in kinematic relations

$$\dot{\mathbf{R}} = \mathbf{R}\widetilde{\boldsymbol{\Omega}}\,, \qquad \dot{\mathbf{x}} = \mathbf{u} \tag{1}$$

with $\mathbf{u} \in \mathbb{R}^3$ denoting the translation velocity in the inertial frame and a skew symmetric matrix

$$\widetilde{\boldsymbol{\Omega}} := \begin{pmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3} \tag{2}$$

that represents the angular velocity $\boldsymbol{\Omega} = (\,\Omega_1,\; \Omega_2,\; \Omega_3\,)^\top \in \mathbb{R}^3$. The semi-direct product $G = \mathrm{SO}(3) \ltimes \mathbb{R}^3$ is known as the *special Euclidean group* SE(3) with the group operation

$$(\mathbf{R}_a,\mathbf{x}_a)\circ(\mathbf{R}_b,\mathbf{x}_b)=(\mathbf{R}_a\mathbf{R}_b,\,\mathbf{R}_a\mathbf{x}_b+\mathbf{x}_a)\,,$$

kinematic relations

$$\dot{\mathbf{R}}=\mathbf{R}\widetilde{\boldsymbol{\Omega}}\,,\qquad \dot{\mathbf{x}}=\mathbf{R}\mathbf{U} \tag{3}$$

and $\mathbf{U}\in\mathbb{R}^3$ denoting the translation velocity in the body-attached frame.

For group elements $q=(\mathbf{R},\mathbf{x})$, the group operations in $\mathrm{SO}(3)\times\mathbb{R}^3$ and in $\mathrm{SE}(3)$ are equivalent to the matrix multiplication of non-singular block-structured matrices in $\mathbb{R}^{7\times7}$ and in $\mathbb{R}^{4\times4}$, respectively, that are defined by

$$\mathrm{SO}(3)\times\mathbb{R}^3\;:\;\begin{pmatrix}\mathbf{R}&\mathbf{0}_{3\times3}&\mathbf{0}_{3\times1}\\ \mathbf{0}_{3\times3}&\mathbf{I}_3&\mathbf{x}\\ \mathbf{0}_{1\times3}&\mathbf{0}_{1\times3}&1\end{pmatrix},\qquad \mathrm{SE}(3)\;:\;\begin{pmatrix}\mathbf{R}&\mathbf{x}\\ \mathbf{0}_{1\times3}&1\end{pmatrix}. \tag{4}$$

Therefore, the groups $\mathrm{SO}(3)\times\mathbb{R}^3$ and $\mathrm{SE}(3)$ as well as the group $\mathrm{SO}(3)$ of all rotation matrices $\mathbf{R}$ are isomorphic to a subset of a general linear group $\mathrm{GL}(r)=\{\,\mathbf{A}\in\mathbb{R}^{r\times r}\,:\,\det\mathbf{A}\neq0\,\}$ of suitable degree $r>0$. The structure of the block matrices in (4) and the orthogonality condition $\mathbf{R}^\top\mathbf{R}=\mathbf{I}_3$ imply that the groups $\mathrm{SO}(3)\times\mathbb{R}^3$, $\mathrm{SE}(3)$ and $\mathrm{SO}(3)$ are isomorphic to differentiable manifolds in $\mathrm{GL}(7)$, $\mathrm{GL}(4)$ and $\mathrm{GL}(3)$, respectively.

2.2 Differential Equations on Manifolds: Matrix Lie Groups

A group G with group operation $\circ$ and neutral element $e\in G$ is called a *Lie group* if G is a differentiable manifold and the group operation $\circ\,:\,G\times G\to G$ as well as the map $q\mapsto q^{-1}$ are differentiable ($q\circ q^{-1}=e$). Lie groups that are subgroups of $\mathrm{GL}(r)$ for some $r>0$ are called *matrix Lie groups* if the group operation $\circ$ is given by the matrix multiplication. For a compact introduction to analytical and numerical aspects of such matrix Lie groups, the interested reader is referred to (Hairer et al. 2006, Sect. IV.6).

It is a trivial observation that a continuously differentiable function $q(t)$ with $q(t_0)\in G$ will remain in a Lie group G if and only if its time derivative $\dot q(t)$ is in the tangent space T_qG at the point $q=q(t)$: $\dot q(t)\in T_{q(t)}G$, $(t\geq t_0)$. The tangent space at the neutral element e defines the *Lie algebra* $\mathfrak{g}:=T_eG$. As a linear space, it is isomorphic to a finite dimensional linear space $\mathbb{R}^k$ with an invertible linear mapping $\widetilde{(\bullet)}\,:\,\mathbb{R}^k\to\mathfrak{g}$, $\mathbf{v}\mapsto\widetilde{\mathbf{v}}$.

The group structure of G makes it possible to represent the elements of T_qG at *any* element $q\in G$ by the elements $\widetilde{\mathbf{v}}$ of the Lie algebra: The left translation

$$L_q\;:\;G\to G\,,\quad y\mapsto L_q(y):=q\circ y$$

defines a bijection in G. Its derivative $DL_q(y)$ at $y = e$ represents the corresponding bijection between the tangent spaces $\mathfrak{g} := T_e G$ and $T_q G$, i.e.,

$$T_q G = \{\, DL_q(e) \cdot \widetilde{\mathbf{v}} \,:\, \widetilde{\mathbf{v}} \in \mathfrak{g} \,\} = \{\, DL_q(e) \cdot \widetilde{\mathbf{v}} \,:\, \mathbf{v} \in \mathbb{R}^k \,\} . \tag{5}$$

With these notations, kinematic relations like (1) and (3) may be summarized in compact form:

$$\dot{q}(t) = DL_{q(t)}(e) \cdot \widetilde{\mathbf{v}}(t) \tag{6}$$

with a velocity vector $\mathbf{v}(t) \in \mathbb{R}^k$. In (6), the left translation L_q as well as the tilde operator $\widetilde{(\bullet)}$ depend on the specific Lie group setting.

For constant velocity $\mathbf{v}$, the kinematic relation (6) yields locally

$$q(t) = q(t_0) \circ \exp\big((t - t_0)\widetilde{\mathbf{v}}\big) \in G \tag{7}$$

with the exponential map $\exp \,:\, \mathfrak{g} \to G$. For matrix Lie groups, this exponential map is given by

$$\exp(\widetilde{\mathbf{v}}) = \sum_{i=0}^{\infty} \frac{1}{i!}\, \widetilde{\mathbf{v}}^i . \tag{8}$$

It is a local diffeomorphism, i.e., for any $q_a \in G$ there are neighbourhoods $U_{q_a} \subset G$ and $V_{\widetilde{\mathbf{0}}} \subset \mathfrak{g}$ such that any $q \in U_{q_a}$ may be expressed by

$$q = q_a \circ \exp(\widetilde{\boldsymbol{\Delta}}_q) \tag{9}$$

with a uniquely defined element $\widetilde{\boldsymbol{\Delta}}_q \in V_{\widetilde{\mathbf{0}}}$.

Example 2.1 (a) Using the block matrix representation (4), the groups $\mathrm{SO}(3) \times \mathbb{R}^3$, SE(3) and SO(3) are seen to be matrix Lie groups. The Lie algebra corresponding to Lie group $G = \mathrm{SO}(3)$ is given by the set

$$\mathfrak{so}(3) := \{\, \mathbf{A} \in \mathbb{R}^{3\times 3} \,:\, \mathbf{A} + \mathbf{A}^\top = \mathbf{0} \,\}$$

of all skew symmetric matrices in $\mathbb{R}^{3\times 3}$. As a linear space, this Lie algebra is isomorphic to $\mathbb{R}^3$ with the tilde operator being defined in (2). In SO(3), the exponential map (8) may be evaluated very efficiently by Rodrigues' formula

$$\exp_{\mathrm{SO}(3)}(\widetilde{\boldsymbol{\Omega}}) = \mathbf{I}_3 + \frac{\sin\Phi}{\Phi}\, \widetilde{\boldsymbol{\Omega}} + \frac{1 - \cos\Phi}{\Phi^2}\, \widetilde{\boldsymbol{\Omega}}^2 \tag{10}$$

with $\Phi := \|\boldsymbol{\Omega}\|_2$ since powers $\widetilde{\boldsymbol{\Omega}}^i$ with $i \geq 3$ may be expressed in terms of $\mathbf{I}_3$, $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^2$ because each matrix $\widetilde{\boldsymbol{\Omega}} \in \mathbb{R}^{3\times 3}$ is a zero of its characteristic polynomial $\chi_\mu(\widetilde{\boldsymbol{\Omega}}) = \det(\mu\mathbf{I}_3 - \widetilde{\boldsymbol{\Omega}}) = \mu^3 + \|\boldsymbol{\Omega}\|_2^2\, \mu = \mu^3 + \Phi^2\mu$, i.e., $\widetilde{\boldsymbol{\Omega}}^3 = -\Phi^2\, \widetilde{\boldsymbol{\Omega}}$ (Cayley-Hamilton theorem).

According to (1), (3) and (6), the Lie algebras of SO(3) × $\mathbb{R}^3$ and SE(3) are parametrized by vectors $\mathbf{v} = (\boldsymbol{\Omega}^\top, \mathbf{u}^\top)^\top$ and $\mathbf{v} = (\boldsymbol{\Omega}^\top, \mathbf{U}^\top)^\top$, respectively. In block matrix form, they are represented by (Brüls et al. 2011)

$$\mathfrak{so}(3) \times \mathbb{R}^3 \,:\, \widetilde{\mathbf{v}} = \begin{pmatrix} \widetilde{\boldsymbol{\Omega}} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{u} \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & 0 \end{pmatrix}, \quad \mathfrak{se}(3) \,:\, \widetilde{\mathbf{v}} = \begin{pmatrix} \widetilde{\boldsymbol{\Omega}} & \mathbf{U} \\ \mathbf{0}_{1\times 3} & 0 \end{pmatrix}$$

with exponential maps

$$\exp_{\mathrm{SO}(3)\times\mathbb{R}^3}(\widetilde{\mathbf{v}}) = \begin{pmatrix} \exp_{\mathrm{SO}(3)}(\widetilde{\boldsymbol{\Omega}}) & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{3\times 3} & \mathbf{I}_3 & \mathbf{u} \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & 1 \end{pmatrix}, \tag{11a}$$

$$\exp_{\mathrm{SE}(3)}(\widetilde{\mathbf{v}}) = \begin{pmatrix} \exp_{\mathrm{SO}(3)}(\widetilde{\boldsymbol{\Omega}}) & \mathbf{T}^\top_{\mathrm{SO}(3)}(\boldsymbol{\Omega})\,\mathbf{U} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix} \tag{11b}$$

and the so-called tangent operator $\mathbf{T}_{\mathrm{SO}(3)} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$, see (33), that will be discussed in more detail in Remark 2.8(b) below.

(b) The linear space $\mathbb{R}^k$ with vector addition + as group operation ∘ is a trivial example of a matrix Lie group since $\mathbf{x} \in \mathbb{R}^k$ may be identified with the non-singular 2 × 2 block matrix

$$\begin{pmatrix} \mathbf{I}_k & \mathbf{x} \\ \mathbf{0}_{1\times k} & 1 \end{pmatrix} \in \mathrm{GL}(k+1)\,. \tag{12}$$

Substituting vector $\mathbf{x}$ by $\mathbf{u} \in \mathbb{R}^k$ and the main diagonal blocks by $\mathbf{0}_{k\times k}$ and by 0, respectively, we get the block matrix representation of the corresponding Lie algebra that is parametrized by $\mathbf{u}$:

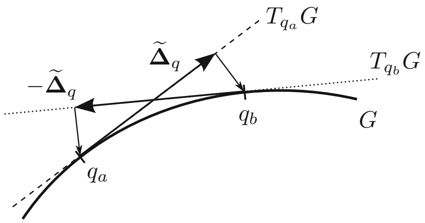
$$\widetilde{\mathbf{u}} = \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{u} \\ \mathbf{0}_{1\times k} & 0 \end{pmatrix}, \quad \exp_{\mathbb{R}^k}(\widetilde{\mathbf{u}}) = \begin{pmatrix} \mathbf{I}_k & \mathbf{u} \\ \mathbf{0}_{1\times k} & 1 \end{pmatrix}. \tag{13}$$

Alternatively, the exponential map may be expressed directly in terms of $\mathbf{u} \in \mathbb{R}^k$ using $\exp_{\mathbb{R}^k} = \mathrm{id}_{\mathbb{R}^k}$, i.e., $\mathbf{x} \circ \exp_{\mathbb{R}^k}(\widetilde{\mathbf{u}}) = \mathbf{x} + \mathbf{u}$.

(c) The block matrix representation of $(\mathbf{R}, \mathbf{x})_{\mathrm{SO}(3)\times\mathbb{R}^3}$ in (4) is block-diagonal with diagonal blocks for $\mathbf{R} \in \mathrm{SO}(3)$ and $\mathbf{x} \in \mathbb{R}^3$, see (12). The same block-diagonal structure is observed for the elements of the corresponding Lie algebra $\mathfrak{so}(3) \times \mathbb{R}^3$, for the tilde operator and for $\exp_{\mathrm{SO}(3)\times\mathbb{R}^3}$, see (11a) and (13). It is typical for direct products of Lie groups and may be used as well for Lie groups $G^N = G \times G \times \cdots \times G$ that are direct products of $N \geq 2$ factors G with $G = \mathrm{SO}(3) \times \mathbb{R}^3$ or $G = \mathrm{SE}(3)$. In particular, we have

$$\exp_{G^N}\big((\widetilde{\mathbf{v}}_1, \widetilde{\mathbf{v}}_2, \ldots, \widetilde{\mathbf{v}}_N)\big) = \operatorname{blockdiag}_{1\leq i\leq N} \exp_G(\widetilde{\mathbf{v}}_i)\,.$$

Fig. 1 Interpolation in Lie groups: $q_b = q_a \circ \exp(\widetilde{\mathbf{\Delta}}_q)$



$T_{q_a}G$
$\widetilde{\mathbf{\Delta}}_q$
$T_{q_b}G$
$-\widetilde{\mathbf{\Delta}}_q$
q_b
G
q_a

Hence, the exponential map $\exp_{G^N} : \mathfrak{g}^N \to G^N$ in the direct product G^N may be evaluated as efficiently as the one in its factors G, see (10) and (11). In flexible multibody dynamics, the configuration spaces $(\mathrm{SO}(3) \times \mathbb{R}^3)^N$ and $(\mathrm{SE}(3))^N$ are of special interest since they allow to represent the configuration of an articulated system of rigid and flexible bodies in the nonlinear finite element method by $N \geq 1$ pairs of absolute nodal translation and rotation variables, see (Brüls et al. 2012; Géradin and Cardona 2001).

Remark 2.2 The parametrization (9) offers a generic way to interpolate between q_a and any point q_b in a sufficiently small neighbourhood $U_{q_a} \subset G$, see Fig. 1: If $q_b = q_a \circ \exp(\widetilde{\mathbf{\Delta}}_q)$ with a vector $\mathbf{\Delta}_q \in \mathbb{R}^k$ of sufficiently small norm $\| \mathbf{\Delta}_q \|$, then $\exp(\vartheta \widetilde{\mathbf{\Delta}}_q)$ is well defined for any $\vartheta \in [0, 1]$ and $q_a, q_b \in G$ are connected by the path

$$\{ \, q(\vartheta; q_a, \mathbf{\Delta}_q) = q_a \circ \exp(\vartheta \widetilde{\mathbf{\Delta}}_q) \; : \; \vartheta \in [0, 1] \, \} \subset G \, .$$

Because of $q_a = q_b \circ \exp(-\widetilde{\mathbf{\Delta}}_q)$ the parametrization of this path by $\vartheta \in [0, 1]$ is symmetric in the sense that $q(\vartheta; q_a, \mathbf{\Delta}_q) = q(1 - \vartheta; q_b, -\mathbf{\Delta}_q)$. This expression is the Lie group equivalent to the identity $\mathbf{q}_a + \vartheta \mathbf{\Delta}_q = \mathbf{q}_b - (1 - \vartheta) \mathbf{\Delta}_q$ that is trivially satisfied for a path that interpolates two points $\mathbf{q}_a, \mathbf{q}_b \in \mathbb{R}^k$.

In the Lie group setting, the *nonlinear* structure of the configuration space G makes it possible to represent large rotations *globally* without singularities. Under reasonable smoothness assumptions, there are smooth functions $q : [t_0, t_{\mathrm{end}}] \to G$ solving the equations of motion on a time interval $[t_0, t_{\mathrm{end}}]$ of finite length, see Sect. 2.3 below. *Locally*, for a fixed time $t = t^* \in [t_0, t_{\mathrm{end}}]$, the configuration space in a sufficiently small neighbourhood of $q(t^*)$ may nevertheless be parametrized by elements of the *linear* space $\mathfrak{g}$ that is independent of t^* and $q(t^*)$, see (9).

The local parametrization of G by elements $\widetilde{\mathbf{v}} \in \mathfrak{g}$ provides the basis for an efficient implementation of Lie group time integration methods and for the analysis of discretization errors, see Sects. 3 and 4 below. Using the notation $\exp(\cdot)$ we will assume tacitly throughout the paper that the argument of the exponential map is in a small neighbourhood of $\widetilde{\mathbf{0}} \in \mathfrak{g}$ on which exp is a diffeomorphism.

The basic concepts of time discretization and error analysis in Lie group time integration are not limited to the specific parametrization by the exponential map, see, e.g., (Kobilarov et al. 2009) for an analysis of variational Lie group integrators that may be combined with the exponential map exp, with the Cayley transform

$\mathrm{cay}(\widetilde{\mathbf{v}}/2) = (\mathbf{I} - \widetilde{\mathbf{v}}/2)^{-1}(\mathbf{I} + \widetilde{\mathbf{v}}/2)$ or with other local parametrizations. In the present paper, we restrict ourselves, however, to the exponential map that reproduces the flow exactly if the velocity $\widetilde{\mathbf{v}} \in \mathfrak{g}$ is constant, see (7).

2.3 Configuration Space with Lie Group Structure: Equations of Motion

In a k-dimensional configuration space G with Lie group structure, the kinematic relations are given by (6) with position coordinates $q(t) \in G$ and the velocity vector $\mathbf{v}(t) \in \mathbb{R}^k$.

We consider constrained systems with $m \leq k$ linearly independent holonomic constraints $\mathbf{\Phi}(q) = \mathbf{0}$ that are coupled by constraint forces $-\mathbf{B}^\top(q)\boldsymbol{\lambda}$ to the equilibrium equations for forces and momenta. Here, $\boldsymbol{\lambda}(t) \in \mathbb{R}^m$ denotes a vector of Lagrange multipliers which is multiplied by the transposed of the constraint matrix $\mathbf{B}(q) \in \mathbb{R}^{m \times k}$ with rank $\mathbf{B}(q) = m$ that represents the constraint gradients in the sense that

$$D\mathbf{\Phi}(q) \cdot \big(DL_q(e) \cdot \widetilde{\mathbf{w}}\big) = \mathbf{B}(q)\mathbf{w}\,, \;\; (\,\mathbf{w} \in \mathbb{R}^k\,)\,. \tag{14}$$

The notation $D\mathbf{\Phi}(q) \cdot \big(DL_q(e) \cdot \widetilde{\mathbf{w}}\big)$ is used for the directional derivative of $\mathbf{\Phi} : G \to \mathbb{R}^m$ at $q \in G$ in the direction of $DL_q(e) \cdot \widetilde{\mathbf{w}} \in T_qG$.

Kinematic equations, equilibrium conditions and holonomic constraints are summarized in the equations of motion

$$\dot{q} = DL_q(e) \cdot \widetilde{\mathbf{v}}\,, \tag{15a}$$

$$\mathbf{M}(q)\dot{\mathbf{v}} = -\mathbf{g}(q, \mathbf{v}, t) - \mathbf{B}^\top(q)\boldsymbol{\lambda}\,, \tag{15b}$$

$$\mathbf{\Phi}(q) = \mathbf{0} \tag{15c}$$

that form a differential-algebraic equation (DAE) on Lie group G, see (Brüls and Cardona 2010). Matrix $\mathbf{M}(q)$ denotes the mass matrix that is supposed to be symmetric, positive definite. The force vector $-\mathbf{g}(q, \mathbf{v}, t)$ summarizes external, internal and complementary inertia forces. Throughout the present paper, we consider equations of motion (15) with functions $\mathbf{M}(q)$, $\mathbf{g}(q, \mathbf{v}, t)$ and $\mathbf{\Phi}(q)$ being smooth in the sense that they are as often continuously differentiable as required by the convergence analysis.

Remark 2.3 (a) For linear configuration spaces, the equations of motion (15) are well known from textbooks on DAE time integration, see, e.g., (Brenan et al. 1996, Sect. 6.2 and Hairer and Wanner 1996, Sect. VII.1). Model equations of constrained mechanical and mechatronic systems in industrial applications have often a more complex structure with additional first-order differential equations $\dot{\mathbf{c}} = \mathbf{h}_\mathbf{c}(q, \mathbf{v}, \mathbf{c}, t)$ or additional algebraic equations $\mathbf{0} = \mathbf{h}_\mathbf{s}(q, \mathbf{s})$ that are locally uniquely solvable w.r.t. $\mathbf{s} = \mathbf{s}(q)$ if the Jacobian $(\partial\mathbf{h}_\mathbf{s}/\partial\mathbf{s})(q, \mathbf{s})$ is non-singular. Other useful generalizations of (15) are rheonomic, i.e., explicitly time-dependent constraints $\mathbf{\Phi}(q,t)=\mathbf{0}$ and force vectors $\mathbf{g}=\mathbf{g}(q,\mathbf{v},\boldsymbol{\lambda},t)$ that contain friction forces depending nonlinearly on $\boldsymbol{\lambda}$, see (Arnold et al. 2011c and Brüls and Golinval 2006) for a more detailed discussion. All these additional model components may be considered straightforwardly in the convergence analysis of generalized-α Lie group integrators, see (Arnold et al. 2015).

(b) The full rank assumption on $\mathbf{B}(q)$ is essential for the analysis and numerical solution of (15) since otherwise the Lagrange multipliers $\boldsymbol{\lambda}(t)$ would not be uniquely defined, see (García de Jalón and Bayo 1994, Sect. 3.4) and the more recent material in (García de Jalón and Gutiérrez-López 2013). On the other hand, the assumptions on $\mathbf{M}(q)$ may be slightly relaxed considering symmetric, positive semi-definite mass matrices that are positive definite on $\ker\mathbf{B}(q)$, see (Géradin and Cardona 2001). The extension of the convergence analysis to this more complex class of model equations has recently been discussed in (Arnold et al. 2014).

The holonomic constraints (15c) imply hidden constraints at the level of velocity coordinates and at the level of acceleration coordinates. The first ones are obtained by differentiation of (15c) w.r.t. t:

$$\mathbf{0}=\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Phi}(q(t))=D\mathbf{\Phi}(q(t))\cdot\dot{q}(t)=D\mathbf{\Phi}(q)\cdot\big(DL_q(e)\cdot\widetilde{\mathbf{v}}\big)=\mathbf{B}(q)\mathbf{v}\,. \tag{16}$$

For the second time derivative of (15c), we have to consider partial derivatives of $\mathbf{\Theta}(q,\mathbf{z}):=\mathbf{B}(q)\mathbf{z}$ w.r.t. $q\in G$. Since $\mathbf{\Theta}\,:\,G\times\mathbb{R}^k\to\mathbb{R}^m$ is by construction linear in $\mathbf{z}$ we have

$$D_q\mathbf{\Theta}(q,\mathbf{z})\cdot\big(DL_q(e)\cdot\widetilde{\mathbf{w}}\big)=\mathbf{Z}(q)(\mathbf{z},\mathbf{w})\,,\quad(\mathbf{w}\in\mathbb{R}^k\,) \tag{17}$$

with a bilinear form $\mathbf{Z}(q)\,:\,\mathbb{R}^k\times\mathbb{R}^k\to\mathbb{R}^m$. Using these notations, the time derivative of (16) gets the form

$$\mathbf{0}=\frac{\mathrm{d}}{\mathrm{d}t}\big(\mathbf{B}(q(t))\mathbf{v}(t)\big)=\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Theta}\big(q(t),\mathbf{v}(t)\big)=\mathbf{B}(q)\dot{\mathbf{v}}+\mathbf{Z}(q)(\mathbf{v},\mathbf{v})\,. \tag{18}$$

It defines the hidden constraints at the level of acceleration coordinates.

The dynamical equations (15b) and the hidden constraints (18) are linear in $\dot{\mathbf{v}}(t)$ and $\boldsymbol{\lambda}(t)$ and may formally be used to eliminate $\boldsymbol{\lambda}(t)$ and to express $\dot{\mathbf{v}}(t)$ in terms of t, $q(t)$ and $\mathbf{v}(t)$, see (Hairer and Wanner 1996, Sect. VII.1):

$$\begin{pmatrix}\mathbf{M}(q) & \mathbf{B}^\top(q)\\ \mathbf{B}(q) & \mathbf{0}\end{pmatrix}\begin{pmatrix}\dot{\mathbf{v}}\\ \boldsymbol{\lambda}\end{pmatrix}=\begin{pmatrix}-\mathbf{g}(q,\mathbf{v},t)\\ -\mathbf{Z}(q)(\mathbf{v},\mathbf{v})\end{pmatrix}. \tag{19}$$

Initial value problems for the resulting analytically equivalent unconstrained system for functions $q\,:\,[t_0,t_{\mathrm{end}}]\to G$ and $\mathbf{v}\,:\,[t_0,t_{\mathrm{end}}]\to\mathbb{R}^k$ are uniquely solvable whenever its right-hand side satisfies a Lipschitz condition, see, e.g., (Walter 1998). This proves unique solvability of initial value problems for the constrained system

(15) if $q(t_0)$ and $\mathbf{v}(t_0)$ are *consistent* with the (hidden) constraints (15c) and (16), i.e., $\mathbf{\Phi}\big(q(t_0)\big) = \mathbf{B}\big(q(t_0)\big)\mathbf{v}(t_0) = \mathbf{0}$. The initial values $\dot{\mathbf{v}}(t_0)$ and $\boldsymbol{\lambda}(t_0)$ are given by (19) with $t = t_0$, $q = q(t_0)$ and $\mathbf{v} = \mathbf{v}(t_0)$.

The index analysis of Lie group DAE (15) follows step by step the classical index analysis for the equations of motion for constrained mechanical systems in linear configuration spaces, see (Hairer and Wanner 1996, Sect. VII.1). The algebraic variables $\boldsymbol{\lambda} = \boldsymbol{\lambda}(q, \mathbf{v}, t)$ are defined by the system of linear equations (19) that contains the second time derivative of (15c). A formal third differentiation step yields $\dot{\boldsymbol{\lambda}} = \dot{\boldsymbol{\lambda}}(q, \mathbf{v}, t)$ and illustrates that (15) is an index-3 Lie group DAE in $G \times \mathbb{R}^k \times \mathbb{R}^m$. Therefore, Eq. (15) is called the *index-3 formulation* of the equations of motion.

Remark 2.4 Block-structured systems of linear equations

$$\begin{pmatrix} \mathbf{M} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{\dot{\mathbf{v}}} \\ \mathbf{x}_{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{\dot{\mathbf{v}}} \\ \mathbf{r}_{\boldsymbol{\lambda}} \end{pmatrix} \tag{20}$$

with a symmetric, positive definite matrix $\mathbf{M} \in \mathbb{R}^{k\times k}$ and a rectangular matrix $\mathbf{B} \in \mathbb{R}^{m\times k}$ of full rank $m \leq k$ are uniquely solvable since left multiplication of the upper block row by $\mathbf{B}\mathbf{M}^{-1}$ yields equations

$$\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^\top \mathbf{x}_{\boldsymbol{\lambda}} = \mathbf{B}\mathbf{M}^{-1}\mathbf{r}_{\dot{\mathbf{v}}} - \mathbf{B}\mathbf{x}_{\dot{\mathbf{v}}} = \mathbf{B}\mathbf{M}^{-1}\mathbf{r}_{\dot{\mathbf{v}}} - \mathbf{r}_{\boldsymbol{\lambda}}$$

that may be solved w.r.t. $\mathbf{x}_{\boldsymbol{\lambda}} \in \mathbb{R}^m$ since $\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^\top$ is symmetric, positive definite. Inserting this vector $\mathbf{x}_{\boldsymbol{\lambda}}$ in the upper block row, we get $\mathbf{x}_{\dot{\mathbf{v}}} \in \mathbb{R}^k$ from $\mathbf{M}\mathbf{x}_{\dot{\mathbf{v}}} = \mathbf{r}_{\dot{\mathbf{v}}} - \mathbf{B}^\top \mathbf{x}_{\boldsymbol{\lambda}}$. The most time-consuming parts of this block Gaussian elimination are the Cholesky factorization of $\mathbf{M} \in \mathbb{R}^{k\times k}$ (to get $\mathbf{M}^{-1}\mathbf{B}^\top \in \mathbb{R}^{k\times m}$ and $\mathbf{M}^{-1}\mathbf{r}_{\dot{\mathbf{v}}} \in \mathbb{R}^k$) the evaluation of the matrix-matrix product $\mathbf{B}(\mathbf{M}^{-1}\mathbf{B}^\top) \in \mathbb{R}^{m\times m}$ and the Cholesky factorization of this matrix.

Alternatively, we could follow a nullspace approach that separates the nullspace of $\mathbf{B} \in \mathbb{R}^{m\times k}$ from a non-singular matrix $\bar{\mathbf{R}} \in \mathbb{R}^{m\times m}$: For any non-singular matrix $\mathbf{Q} \in \mathbb{R}^{k\times k}$ with $\mathbf{B}\mathbf{Q} = \big(\bar{\mathbf{R}}^\top, \ \mathbf{0}_{m\times(k-m)}\big)$, system (20) is equivalent to

$$\begin{pmatrix} \bar{\mathbf{M}}_{11} & \bar{\mathbf{M}}_{12} & \bar{\mathbf{R}} \\ \bar{\mathbf{M}}_{21} & \bar{\mathbf{M}}_{22} & \mathbf{0} \\ \bar{\mathbf{R}}^\top & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{x}}_{\dot{\mathbf{v}},1} \\ \bar{\mathbf{x}}_{\dot{\mathbf{v}},2} \\ \mathbf{x}_{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{r}}_{\dot{\mathbf{v}},1} \\ \bar{\mathbf{r}}_{\dot{\mathbf{v}},2} \\ \mathbf{r}_{\boldsymbol{\lambda}} \end{pmatrix} \text{ with } \begin{pmatrix} \bar{\mathbf{M}}_{11} & \bar{\mathbf{M}}_{12} \\ \bar{\mathbf{M}}_{21} & \bar{\mathbf{M}}_{22} \end{pmatrix} = \mathbf{Q}^\top \mathbf{M}\mathbf{Q}\,,$$

$\bar{\mathbf{x}}_{\dot{\mathbf{v}}} = \mathbf{Q}^{-1}\mathbf{x}_{\dot{\mathbf{v}}}$ and $\bar{\mathbf{r}}_{\dot{\mathbf{v}}} = \mathbf{Q}^\top \mathbf{r}_{\dot{\mathbf{v}}}$. This block-structured system may be solved in three steps by block backward substitution to get $\bar{\mathbf{x}}_{\dot{\mathbf{v}},1}$, $\bar{\mathbf{x}}_{\dot{\mathbf{v}},2}$ and $\bar{\mathbf{x}}_{\boldsymbol{\lambda}}$ since matrices $\bar{\mathbf{R}}^\top$, $\bar{\mathbf{M}}_{22}$ and $\bar{\mathbf{R}}$ are non-singular. Betsch and Leyendecker (2006) discussed analytical nullspace representations of the constraint matrix $\mathbf{B}$ for typical types of constraints in engineering systems. If such analytical expressions are not available, then matrices $\mathbf{Q}$ and $\bar{\mathbf{R}}$ could be computed, e.g., by a QR-factorization of $\mathbf{B}^\top \in \mathbb{R}^{k\times m}$, see (Golub and van Loan 1996).

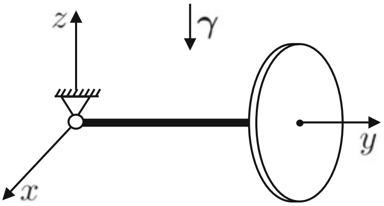


Fig. 2 Benchmark problem Heavy top (Brüls and Cardona 2010), see also (Géradin and Cardona 2001)

2.4 Benchmark Problem: Heavy Top

The Lie group formulation of the equations of motion is the backbone of a rather general finite element framework for flexible multibody dynamics (Géradin and Cardona 2001). In the present paper, we focus on basic aspects of Lie group time integration in multibody dynamics and restrict the numerical tests to the simulation of a single rigid body in a gravitation field. This *heavy top* has found much interest in mechanics and serves as a benchmark problem for Lie group methods (Géradin and Cardona 2001, Sect. 5.8). The simulation of more complex flexible structures by Lie group time integration methods is discussed, e.g., in (Brüls et al. 2012).

Figure 2 shows the configuration of the heavy top in $\mathbb{R}^3$ with $\mathbf{R}(t) \in \mathrm{SO}(3)$ characterizing its orientation and the position vector $\mathbf{x}(t) \in \mathbb{R}^3$ of the centre of mass in the inertial frame. In the body-attached frame, the centre of mass is given by $\mathbf{X} = (\,0,\ 1,\ 0\,)^\top$. Here and in the following, we omit all physical units. We consider a gravitation field with fixed acceleration vector $\boldsymbol{\gamma} = (0, 0, -9.81)^\top$. Mass and inertia tensor are given by $m = 15.0$ and $\mathbf{J} = \mathrm{diag}\,(0.234375, 0.46875, 0.234375)$ with $\mathbf{J}$ denoting the inertia tensor w.r.t. the centre of mass.

In the benchmark problem, the top rotates about a fixed point. Therefore, the configuration variables $(\mathbf{R}, \mathbf{x})$ are subject to holonomic constraints $\mathbf{x} = \mathbf{R}\mathbf{X}$. We consider an initial configuration being defined by $\mathbf{R}(0) = \mathbf{I}_3$ with an angular velocity $\boldsymbol{\Omega}(0) = (0, 150, -4.61538)^\top$. All other initial values are supposed to be consistent with $\mathbf{0} = \boldsymbol{\Phi}\big((\mathbf{R}, \mathbf{x})\big) := \mathbf{X} - \mathbf{R}^\top\mathbf{x}$ and with the corresponding hidden constraints (16) and (18) at the level of velocity and acceleration coordinates.

The equations of motion (15) of the rotating heavy top result from the principles of classical mechanics. In (Brüls et al. 2011), they were derived for configuration spaces $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and $G = \mathrm{SE}(3)$ following an augmented Lagrangian method. In $\mathrm{SO}(3) \times \mathbb{R}^3$, we get hidden constraints

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{X} - \mathbf{R}^\top\mathbf{x}) = -\dot{\mathbf{R}}^\top\mathbf{x} - \mathbf{R}^\top\dot{\mathbf{x}} = -\widetilde{\boldsymbol{\Omega}}^\top\mathbf{R}^\top\mathbf{x} - \mathbf{R}^\top\mathbf{u} = -\widetilde{\mathbf{X}}\boldsymbol{\Omega} - \mathbf{R}^\top\mathbf{u}$$

and a constraint matrix $\mathbf{B} = (-\widetilde{\mathbf{X}} \quad -\mathbf{R}^\top)$. The equations of motion are given by

$$\mathbf{J}\dot{\boldsymbol{\Omega}} + \boldsymbol{\Omega} \times \mathbf{J}\boldsymbol{\Omega} + \mathbf{X} \times \boldsymbol{\lambda} = \mathbf{0}\,, \tag{21a}$$

$$m\dot{\mathbf{u}} - \mathbf{R}\boldsymbol{\lambda} = m\boldsymbol{\gamma}\,, \tag{21b}$$

$$\mathbf{X} - \mathbf{R}^\top\mathbf{x} = \mathbf{0} \tag{21c}$$

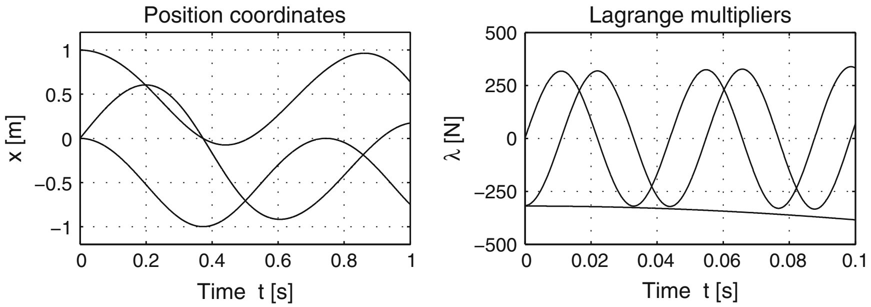


Fig. 3 Heavy top benchmark, $G = \mathrm{SO}(3) \times \mathbb{R}^3$: Reference solution

with kinematic relations (1). Figure 3 shows a reference solution that has been computed with the very small time step size $h = 2.5 \times 10^{-5}$. The position $\mathbf{x}(t) \in \mathbb{R}^3$ of the centre of mass varies slowly in the inertial frame. For the Lagrange multipliers $\boldsymbol{\lambda}(t) \in \mathbb{R}^3$, we observe much higher frequencies that reflect the fast rotation of the top being caused by the rather large initial velocity $\boldsymbol{\Omega}(0)$. Note, that the time scale in the right plot of Fig. 3 has been zoomed by a factor of 10.

In the configuration space SE(3), we have $\dot{\mathbf{x}} = \mathbf{RU}$ resulting in hidden constraints $\mathbf{0} = -\widetilde{\mathbf{X}}\boldsymbol{\Omega} - \mathbf{U}$ with a constraint matrix $\mathbf{B} = (-\widetilde{\mathbf{X}} \quad -\mathbf{I}_3)$ that is constant and does not depend on $q \in G$. The equations of motion are given by

$$\mathbf{J}\dot{\boldsymbol{\Omega}} + \boldsymbol{\Omega} \times \mathbf{J}\boldsymbol{\Omega} + \mathbf{X} \times \boldsymbol{\lambda} = \mathbf{0}\,, \tag{22a}$$

$$m\dot{\mathbf{U}} + m\boldsymbol{\Omega} \times \mathbf{U} - \boldsymbol{\lambda} = \mathbf{R}^\top m\boldsymbol{\gamma}\,, \tag{22b}$$

$$\mathbf{X} - \mathbf{R}^\top\mathbf{x} = \mathbf{0} \tag{22c}$$

with kinematic relations (3). The position coordinates $q = (\mathbf{R}, \mathbf{x})$ coincide for both formulations (21) and (22) but there may be substantial differences between the velocity coordinates $\mathbf{u}(t)$ in the inertial frame and their counterparts $\mathbf{U}(t)$ in the body-attached frame. This is illustrated by the simulation results in Fig. 4 that have been obtained again with time step size $h = 2.5 \times 10^{-5}$. In $\mathrm{SO}(3) \times \mathbb{R}^3$, we observe low frequency changes of $\mathbf{u}(t)$ that correspond to the solution behaviour of $\mathbf{x}(t)$ in the left plot of Fig. 3. For the configuration space $G = \mathrm{SE}(3)$, we see in the right plot of Fig. 4 the dominating influence of the large initial velocity $\boldsymbol{\Omega}(0)$ on the qualitative solution behaviour of $\mathbf{U}(t)$.

Throughout the paper, we will use the two different formulations (21) and (22) of the heavy top benchmark problem for numerical tests to discuss various aspects of the convergence analysis for the generalized-α Lie group integrator.

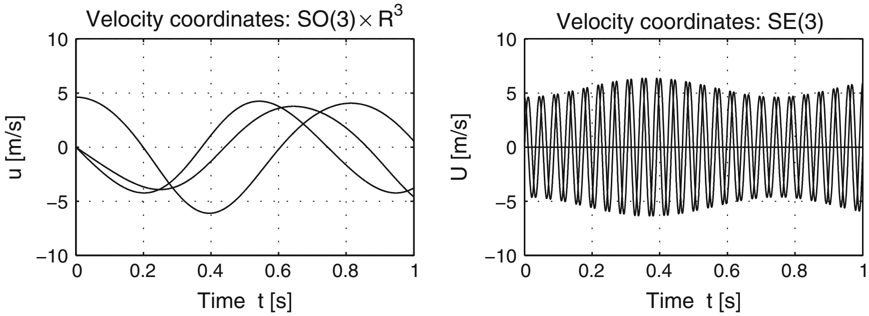


Fig. 4 Heavy top benchmark: Velocity coordinates in the inertial frame ($\mathbf{u}(t)$, *left plot*) and in the body-attached frame ($\mathbf{U}(t)$, *right plot*)

2.5 More on the Exponential Map

Equation (7) illustrates the crucial role of the exponential map for multibody system models that have a configuration space with Lie group structure. Since the numerical solution proceeds in time steps, we have to study the composition of exponential maps with different arguments in more detail. Furthermore, the proposed Lie group time integration methods are implicit and rely on a Newton–Raphson iteration that requires the efficient evaluation of Jacobians $(\partial\mathbf{h}/\partial\mathbf{v})\big(q \circ \exp(\widetilde{\mathbf{v}})\big)$ for vector-valued functions $\mathbf{h} : G \to \mathbb{R}^l$. In the present section, we follow the presentation in (Hairer et al. 2006, Sect. III.4) to discuss these rather technical aspects of Lie group time integration.

For matrix Lie groups, the exponential map exp is given by the matrix exponential. For $s \in \mathbb{R}$ and any matrices $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{r\times r}$, the series expansion (8) shows

$$\begin{aligned}\exp(s\mathbf{A})\exp(s\mathbf{C}) &= (\mathbf{I}_r + s\mathbf{A} + \frac{s^2}{2}\mathbf{A}^2)(\mathbf{I}_r + s\mathbf{C} + \frac{s^2}{2}\mathbf{C}^2) + \mathcal{O}(s^3)\\ &= \mathbf{I}_r + s(\mathbf{A}+\mathbf{C}) + \frac{s^2}{2}(\mathbf{A}^2 + 2\mathbf{A}\mathbf{C} + \mathbf{C}^2) + \mathcal{O}(s^3)\\ &= \mathbf{I}_r + s(\mathbf{A}+\mathbf{C}) + \frac{s^2}{2}(\mathbf{A}+\mathbf{C})^2 + \frac{1}{2}[s\mathbf{A}, s\mathbf{C}] + \mathcal{O}(s^3)\\ &= \exp\big(s\mathbf{A} + s\mathbf{C} + \frac{1}{2}[s\mathbf{A}, s\mathbf{C}]\big) + \mathcal{O}(s^3)\,,\ \ (s \to 0)\end{aligned}$$

with the *matrix commutator* $[\mathbf{A}, \mathbf{C}] := \mathbf{A}\mathbf{C} - \mathbf{C}\mathbf{A}$ that vanishes iff matrices $\mathbf{A}$ and $\mathbf{C}$ commute. For a slightly more detailed analysis of the product of matrix exponentials, we use the Baker–Campbell–Hausdorff formula, see (Hairer et al. 2006, Lemma III.4.3), to get the following estimate:

Lemma 2.5 *For $s \to 0$, the product of matrix exponentials* $\exp(s\mathbf{A})$ *and* $\exp(s\mathbf{C})$ *satisfies*

$$\exp(s\mathbf{A})\exp(s\mathbf{C}) = \exp\bigl(s\mathbf{A} + s\mathbf{C} + \frac{1}{2}[s\mathbf{A}, s\mathbf{C}] + \mathcal{O}(s)\|[s\mathbf{A}, s\mathbf{C}]\|\bigr)\,. \qquad (23)$$

Proof The Baker–Campbell–Hausdorff formula defines the argument of the matrix exponential at the right-hand side of (23) by the solution of an initial value problem with zero initial values at $s = 0$. Solving this initial value problem by Picard iteration with starting guess $s\mathbf{A} + s\mathbf{C} + [s\mathbf{A}, s\mathbf{C}]/2$, we may show that all higher order terms result in a remainder term of size $\mathcal{O}(s)\|[s\mathbf{A}, s\mathbf{C}]\|$, see (23). □

For fixed argument $\mathbf{A}$, the matrix commutator defines a linear operator

$$\mathrm{ad}_{\mathbf{A}} \,:\, \mathbb{R}^{r\times r} \to \mathbb{R}^{r\times r}\,,\ \mathbf{C} \mapsto \mathrm{ad}_{A} := [\mathbf{A}, \mathbf{C}] \qquad (24)$$

that is called the *adjoint operator*. By recursive application of $\mathrm{ad}_{\mathbf{A}}$ we may represent directional derivatives of the exponential map $\exp(\mathbf{A}) = \sum_i \mathbf{A}^i / i!$ in compact form: We denote $\mathrm{ad}^0_{\mathbf{A}}(\mathbf{C}) := \mathbf{C}$ and

$$\mathrm{ad}^{j+1}_{\mathbf{A}}(\mathbf{C}) := \mathrm{ad}_{\mathbf{A}}\bigl(\mathrm{ad}^{j}_{\mathbf{A}}(\mathbf{C})\bigr) = \mathbf{A}\,\mathrm{ad}^{j}_{\mathbf{A}}(\mathbf{C}) - \mathrm{ad}^{j}_{\mathbf{A}}(\mathbf{C})\,\mathbf{A}\,,\ (\, j \geq 1\,) \qquad (25)$$

and consider powers $(\mathbf{A} + s\mathbf{C})^i$, $(\, i \geq 0\,)$, in the limit case $s \to 0$. For $i = 2$, we get

$$(\mathbf{A} + s\mathbf{C})^2 = \mathbf{A}^2 + s(\mathbf{AC} + \mathbf{CA}) + \mathcal{O}(s^2) = \mathbf{A}^2 + s\bigl(2\mathbf{AC} + \mathrm{ad}_{-\mathbf{A}}(\mathbf{C})\bigr) + \mathcal{O}(s^2)\,.$$

Here, the term $\mathrm{ad}_{-\mathbf{A}}(\mathbf{C})$ results from the non-commutativity of matrix multiplication and could be represented as well by the adjoint operator $\mathrm{ad}_{\mathbf{A}}$ itself since $2\mathbf{AC} + \mathrm{ad}_{-\mathbf{A}}(\mathbf{C}) = 2\mathbf{CA} + \mathrm{ad}_{\mathbf{A}}(\mathbf{C})$, see (Hairer et al. 2006). The use of $\mathrm{ad}_{-\mathbf{A}}$ corresponds, however, to the characterization of the tangent space $T_q G$ by left translations L_q, see (5) and the discussion in (Iserles et al. 2000). In multibody dynamics, this characterization implies that vector $\mathbf{v}$ in the kinematic relations (6) is a left-invariant velocity vector. These left-invariant vectors are favourable since the associated rotational inertia are defined in the body-attached frame and the body mass matrices remain constant during motion (Brüls et al. 2011).

Lemma 2.6 *For $s \to 0$ and matrices $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{r\times r}$, the asymptotic behaviour of* $(\mathbf{A} + s\mathbf{C})^i$ *and* $\exp(\mathbf{A} + s\mathbf{C})$ *is characterized by*

$$(\mathbf{A} + s\mathbf{C})^i = \mathbf{A}^i + s\sum_{j=0}^{i-1}\binom{i}{j+1}\mathbf{A}^{i-j-1}\,\mathrm{ad}^{j}_{-\mathbf{A}}(\mathbf{C}) + \mathcal{O}(s^2)\,,\ (\, i \geq 1\,)\,, \qquad (26)$$

and

$$\exp(\mathbf{A} + s\mathbf{C}) = \exp(\mathbf{A})\,\bigl(\mathbf{I}_r + s\ \mathrm{dexp}_{-\mathbf{A}}(\mathbf{C})\bigr) + \mathcal{O}(s^2) \qquad (27)$$

with the matrix-valued function

$$\mathrm{dexp}_{-\mathbf{A}}(\mathbf{C}) := \sum_{j=0}^{\infty} \frac{1}{(j+1)!}\ \mathrm{ad}_{-\mathbf{A}}^{j}(\mathbf{C}) \tag{28}$$

that satisfies $\mathrm{dexp}_{-\mathbf{A}}(\mathbf{C}) = \mathbf{C}$ *whenever* $\mathbf{A}$ *and* $\mathbf{C}$ *commute.*

Proof To prove (26) by induction, we multiply this expression from the right by $(\mathbf{A} + s\mathbf{C})$ and observe that $\mathbf{A}^{i}\, s\mathbf{C} = s\mathbf{A}^{(i+1)-j-1}\,\mathrm{ad}_{-\mathbf{A}}^{j}(\mathbf{C})$ with $j = 0$. Taking into account the identity

$$\mathbf{A}^{i-j-1}\,\mathrm{ad}_{-\mathbf{A}}^{j}(\mathbf{C})\,\mathbf{A} = \mathbf{A}^{(i+1)-(j+1)-1}\,\mathrm{ad}_{-\mathbf{A}}^{j+1}(\mathbf{C}) + \mathbf{A}^{(i+1)-j-1}\,\mathrm{ad}_{-\mathbf{A}}^{j}(\mathbf{C})\,,$$

see (25), we get (26) with i being substituted by $i+1$ since

$$\binom{i}{j} + \binom{i}{j+1} = \binom{i+1}{j+1}, \ \ (\, j = 0, 1, \ldots, i-1\,)\,.$$

For the proof of (27), we scale (26) by $1/i!$ and use the series expansion (8) to get

$$\begin{aligned}
\exp(\mathbf{A} + s\mathbf{C}) &= \sum_{i=0}^{\infty} \frac{1}{i!}\mathbf{A}^{i} + s\sum_{i=0}^{\infty} \frac{1}{i!} \sum_{j=0}^{i-1} \binom{i}{j+1} \mathbf{A}^{i-j-1}\ \mathrm{ad}_{-\mathbf{A}}^{j}(\mathbf{C}) + \mathcal{O}(s^2) \\
&= \sum_{i=0}^{\infty} \frac{1}{i!}\mathbf{A}^{i} + s\sum_{j=0}^{\infty} \frac{1}{(j+1)!} \underbrace{\sum_{i=j+1}^{\infty} \frac{1}{(i-j-1)!}\mathbf{A}^{i-j-1}}_{= \sum_{i=0}^{\infty} \frac{1}{i!}\mathbf{A}^{i} = \exp(\mathbf{A})}\ \mathrm{ad}_{-\mathbf{A}}^{j}(\mathbf{C}) + \mathcal{O}(s^2) \\
&= \exp(\mathbf{A})\,\big(\mathbf{I}_r + s\,\mathrm{dexp}_{-\mathbf{A}}(\mathbf{C})\big) + \mathcal{O}(s^2)\,.
\end{aligned}$$

For commuting matrices $\mathbf{A}$ and $\mathbf{C}$, the iterated adjoint operators $\mathrm{ad}_{-\mathbf{A}}^{j}(\mathbf{C})$ vanish for all $j > 0$ resulting in $\mathrm{dexp}_{-\mathbf{A}}(\mathbf{C}) = \mathbf{C}$, see (28). □

Lemma 2.6 shows that the directional derivative of the matrix exponential is given by $(\partial/\partial\mathbf{A})\exp(\mathbf{A})\mathbf{C} = \exp(\mathbf{A})\,\mathrm{dexp}_{-\mathbf{A}}(\mathbf{C})$. In the Lie group setting, we use this expression to study the Jacobian of vector-valued functions $\mathbf{h}\big(q \circ \exp(\widetilde{\mathbf{v}})\big)$ w.r.t. $\mathbf{v} \in \mathbb{R}^k$. For elements $\widetilde{\mathbf{v}}, \widetilde{\mathbf{w}} \in \mathfrak{g}$, the terms $\mathrm{ad}_{-\widetilde{\mathbf{v}}}(\widetilde{\mathbf{w}})$ and $\mathrm{dexp}_{-\widetilde{\mathbf{v}}}(\widetilde{\mathbf{w}})$ are linear in $\mathbf{w} \in \mathbb{R}^k$ and may be represented by matrix-vector products in $\mathbb{R}^k$ using the notation

$$\widehat{(\bullet)}\ :\ \mathbb{R}^k \to \mathbb{R}^{k\times k} \quad \text{with} \quad \widetilde{\widehat{\mathbf{v}}\mathbf{w}} = \mathrm{ad}_{\widetilde{\mathbf{v}}}(\widetilde{\mathbf{w}}) = [\widetilde{\mathbf{v}}, \widetilde{\mathbf{w}}]\,,\ (\mathbf{v}, \mathbf{w} \in \mathbb{R}^k\,)\,. \tag{29}$$

With (29), the operators $\mathrm{ad}_{\widetilde{\mathbf{v}}}$, $\mathrm{ad}_{-\widetilde{\mathbf{v}}}$ and $\mathrm{ad}_{-\widetilde{\mathbf{v}}}^{j}$ correspond to $k \times k$-matrices $\widehat{\mathbf{v}}$, $-\widehat{\mathbf{v}}$ and $(-\widehat{\mathbf{v}})^{j}$, respectively, and the counterpart to $\widetilde{\mathbf{z}} = \mathrm{dexp}_{-\widetilde{\mathbf{v}}}(\widetilde{\mathbf{w}}) \in \mathfrak{g}$, see (28), is given by $\mathbf{z} = \mathbf{T}(\mathbf{v})\mathbf{w} \in \mathbb{R}^k$ with the *tangent operator*

$$\mathbf{T} : \mathbb{R}^k \to \mathbb{R}^{k\times k}\,, \quad \mathbf{T}(\mathbf{v}) = \sum_{i=0}^{\infty} \frac{(-1)^i}{(i+1)!}\,\widehat{\mathbf{v}}^i\,, \tag{30}$$

see (Iserles et al. 2000). Using the chain rule, we obtain

Corollary 2.7 *Consider a continuously differentiable function* $\mathbf{h}\colon G \to \mathbb{R}^l$ *and a matrix-valued function* $\mathbf{H}\colon G \to \mathbb{R}^{l\times k}$ *that represents the derivative of* $\mathbf{h}$ *in the sense that*

$$D\mathbf{h}(q)\cdot\big(DL_q(e)\cdot\widetilde{\mathbf{w}}\big) = \mathbf{H}(q)\mathbf{w}\,,\ \ (\,\mathbf{w}\in\mathbb{R}^k\,)\,,$$

see (14). The Jacobian of $\mathbf{h}\big(q\circ\exp(\widetilde{\mathbf{v}})\big)$ *w.r.t.* $\mathbf{v}\in\mathbb{R}^k$ *is given by*

$$\frac{\partial\mathbf{h}}{\partial\mathbf{v}}\big(q\circ\exp(\widetilde{\mathbf{v}})\big) = \mathbf{H}\big(q\circ\exp(\widetilde{\mathbf{v}})\big)\mathbf{T}(\mathbf{v})\,. \tag{31}$$

Remark 2.8 (a) For commuting elements of the Lie algebra ($\widetilde{\mathbf{v}}, \widetilde{\mathbf{w}}\in\mathfrak{g}$ with $[\widetilde{\mathbf{v}},\widetilde{\mathbf{w}}] = \widetilde{\mathbf{0}}$), the adjoint operator vanishes resulting in $\widehat{\mathbf{v}}\mathbf{w} = \mathbf{0}_k$ and $\mathbf{T}(\mathbf{v})\mathbf{w} = \mathbf{w}$. Therefore, the tangent operator satisfies $\mathbf{T}(\mathbf{v})\mathbf{v} = \mathbf{v}$, ($\mathbf{v}\in\mathbb{R}^k$), and Corollary 2.7 implies

$$\frac{\mathrm{d}\mathbf{h}}{\mathrm{d}\vartheta}\big(q\circ\exp(\vartheta\widetilde{\mathbf{v}})\big) = \mathbf{H}\big(q\circ\exp(\vartheta\widetilde{\mathbf{v}})\big)\mathbf{v} \tag{32}$$

with $\vartheta\in\mathbb{R}$ and any vector $\mathbf{v}\in\mathbb{R}^k$.

(b) The efficient evaluation of the tangent operator is essential for an efficient implementation of implicit Lie group integrators. In the Lie group $G = \mathrm{SO}(3)$, the hat operator maps $\boldsymbol{\Omega}\in\mathbb{R}^3$ to $\widehat{\boldsymbol{\Omega}} := \widetilde{\boldsymbol{\Omega}}$ with the skew symmetric matrix $\widetilde{\boldsymbol{\Omega}}$ being defined in (2). Similar to Rodrigues' formula (10), the tangent operator $\mathbf{T}_{\mathrm{SO}(3)}$ may be evaluated in closed form (Brüls et al. 2011):

$$\mathbf{T}_{\mathrm{SO}(3)}(\boldsymbol{\Omega}) = \mathbf{I}_3 + \frac{\cos\Phi - 1}{\Phi^2}\,\widetilde{\boldsymbol{\Omega}} + \frac{1-\dfrac{\sin\Phi}{\Phi}}{\Phi^2}\,\widetilde{\boldsymbol{\Omega}}^2\,. \tag{33}$$

For $G = \mathrm{SO}(3)\times\mathbb{R}^3$, the Lie algebra $\mathfrak{g} = \mathfrak{so}(3)\times\mathbb{R}^3$ is parametrized by vectors $\mathbf{v} = (\boldsymbol{\Omega}^\top, \mathbf{u}^\top)^\top\in\mathbb{R}^6$ and we get

$$\widehat{\mathbf{v}} = \mathrm{blockdiag}\,(\,\widetilde{\boldsymbol{\Omega}},\ \mathbf{0}_{3\times3}\,)\,,\quad \mathbf{T}_{\mathrm{SO}(3)\times\mathbb{R}^3}(\mathbf{v}) = \mathrm{blockdiag}\,\big(\,\mathbf{T}_{\mathrm{SO}(3)}(\boldsymbol{\Omega}),\ \mathbf{I}_3\,\big)\,.$$

More complex expressions are obtained for the Lie group $G = \mathrm{SE}(3)$ and its Lie algebra $\mathfrak{se}(3)$ that is parametrized by vectors $\mathbf{v} = (\boldsymbol{\Omega}^\top, \mathbf{U}^\top)^\top\in\mathbb{R}^6$ with

$$\widehat{\mathbf{v}} = \begin{pmatrix} \widetilde{\boldsymbol{\Omega}} & \mathbf{0}_{3\times3} \\ \widetilde{\mathbf{U}} & \widetilde{\boldsymbol{\Omega}} \end{pmatrix}. \tag{34}$$

Using the identities $\widetilde{\boldsymbol{\Omega}}^3 = -\Phi^2\widetilde{\boldsymbol{\Omega}}$, $\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Omega}} = -(\boldsymbol{\Omega}^\top\mathbf{U})\mathbf{I}_3 + \boldsymbol{\Omega}\mathbf{U}^\top$, $\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Omega}}^2 + \widetilde{\boldsymbol{\Omega}}^2\widetilde{\mathbf{U}} = -\Phi^2\widetilde{\mathbf{U}} - (\boldsymbol{\Omega}^\top\mathbf{U})\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Omega}} = -(\boldsymbol{\Omega}^\top\mathbf{U})\widetilde{\boldsymbol{\Omega}}$ with $\Phi := \|\boldsymbol{\Omega}\|_2$, we prove by induction

$$\widehat{\mathbf{v}}^{2l+1} = \begin{pmatrix} (-\Phi^2)^l\,\widetilde{\boldsymbol{\Omega}} & \mathbf{0}_{3\times 3} \\ (-\Phi^2)^l\,\widetilde{\mathbf{U}} - 2l(-\Phi^2)^{l-1}(\boldsymbol{\Omega}^\top\mathbf{U})\widetilde{\boldsymbol{\Omega}} & (-\Phi^2)^l\,\widetilde{\boldsymbol{\Omega}} \end{pmatrix}$$

and

$$\widehat{\mathbf{v}}^{2l+2} = \begin{pmatrix} (-\Phi^2)^l\,\widetilde{\boldsymbol{\Omega}}^2 & \mathbf{0}_{3\times 3} \\ (-\Phi^2)^l\,(\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Omega}} + \widetilde{\boldsymbol{\Omega}}\widetilde{\mathbf{U}}) - 2l(-\Phi^2)^{l-1}(\boldsymbol{\Omega}^\top\mathbf{U})\widetilde{\boldsymbol{\Omega}}^2 & (-\Phi^2)^l\,\widetilde{\boldsymbol{\Omega}}^2 \end{pmatrix}$$

for all $l \geq 0$ and get the tangent operator

$$\mathbf{T}_{\mathrm{SE}(3)}(\boldsymbol{\Omega}) = \begin{pmatrix} \mathbf{T}_{\mathrm{SO}(3)}(\boldsymbol{\Omega}) & \mathbf{0}_{3\times 3} \\ \mathbf{S}_{\mathrm{SE}(3)}(\boldsymbol{\Omega}, \mathbf{U}) & \mathbf{T}_{\mathrm{SO}(3)}(\boldsymbol{\Omega}) \end{pmatrix} \tag{35}$$

with $\mathbf{S}_{\mathrm{SE}(3)}(\mathbf{0}, \mathbf{U}) = -\widetilde{\mathbf{U}}/2$ and

$$\begin{aligned} \mathbf{S}_{\mathrm{SE}(3)}(\boldsymbol{\Omega}, \mathbf{U}) = \frac{1}{\Phi^2}\Big(&-(1-\cos\Phi)\widetilde{\mathbf{U}} + \big(1 - \frac{\sin\Phi}{\Phi}\big)(\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Omega}} + \widetilde{\boldsymbol{\Omega}}\widetilde{\mathbf{U}}) + \\ &+ \big(2\,\frac{1-\cos\Phi}{\Phi^2} - \frac{\sin\Phi}{\Phi}\big)(\boldsymbol{\Omega}^\top\mathbf{U})\widetilde{\boldsymbol{\Omega}} + \\ &+ \frac{1}{\Phi^2}\big(1 - \cos\Phi - 3\,(1 - \frac{\sin\Phi}{\Phi})\big)(\boldsymbol{\Omega}^\top\mathbf{U})\widetilde{\boldsymbol{\Omega}}^2\Big) \end{aligned}$$

if $\boldsymbol{\Omega} \neq \mathbf{0}$, see (Brüls et al. 2011 and Sonneville et al. 2014, Appendix A).

(c) If $\mathbb{R}^k$ with the addition is considered as a Lie group, then we get $\widehat{\mathbf{v}} = \mathbf{0}_{k\times k}$ and $\mathbf{T}_{\mathbb{R}^k}(\mathbf{v}) = \mathbf{I}_k$ for any vector $\mathbf{v} \in \mathbb{R}^k$ since the group operation is commutative.

(d) Similar to the discussion in Example 2.1(c), we observe for direct products like $\mathrm{SO}(3) \times \mathbb{R}^3$ that the matrix $\widehat{\mathbf{v}}$ and the tangent operator $\mathbf{T}(\mathbf{v})$ are block-diagonal. In $(\mathrm{SO}(3) \times \mathbb{R}^3)^N$ and $(\mathrm{SE}(3))^N$, the tangent operators are given by

$$\mathbf{T}_{G^N}\big((\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)\big) = \mathrm{blockdiag}_{1\leq i\leq N}\,\mathbf{T}_G(\mathbf{v}_i) \in \mathbb{R}^{6N\times 6N}$$

with $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and $G = \mathrm{SE}(3)$, respectively.

3 Generalized-α Lie Group Time Integration

The time integration of the equations of motion (15) by Lie group methods is based on the observation that (15a) implies

$$q(t+h) = q(t) \circ \exp\big(h\widetilde{\mathbf{v}}(t) + \frac{h^2}{2}\dot{\widetilde{\mathbf{v}}}(t) + \mathcal{O}(h^3)\big)\,, \quad (\,h \to 0\,)\,. \tag{36}$$

In Sect. 3.1, a generalized-α Lie group method for the index-3 formulation (15) is introduced. In Sect. 3.2, we recall some well-known facts about order, stability and "overshooting" of generalized-α methods in linear spaces. For the heavy top benchmark problem, second-order convergence of the Lie group integrator and an order reduction phenomenon in the transient phase may be observed numerically (Sect. 3.3). In Sect. 3.4, we show that the error constant of the first-order error term may be reduced drastically by an analytical index reduction before time discretization. Implementation aspects and the discretization errors in hidden constraints are studied in Sects. 3.5 and 3.6.

3.1 The Lie Group Time Integration Method

As proposed by Brüls and Cardona (2010), we consider a generalized-α method for the index-3 formulation (15) of the equations of motion that updates the numerical solution $(q_n, \mathbf{v}_n, \mathbf{a}_n, \boldsymbol{\lambda}_n)$ in a time step $t_n \rightarrow t_n + h$ of step size h according to

$$q_{n+1} = q_n \circ \exp(h\widetilde{\boldsymbol{\Delta}\mathbf{q}}_n)\,, \tag{37a}$$

$$\boldsymbol{\Delta}\mathbf{q}_n = \mathbf{v}_n + (0.5-\beta)h\mathbf{a}_n + \beta h\mathbf{a}_{n+1}\,, \tag{37b}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + (1-\gamma)h\mathbf{a}_n + \gamma h\mathbf{a}_{n+1}\,, \tag{37c}$$

$$(1-\alpha_m)\mathbf{a}_{n+1} + \alpha_m\mathbf{a}_n = (1-\alpha_f)\dot{\mathbf{v}}_{n+1} + \alpha_f\dot{\mathbf{v}}_n \tag{37d}$$

with vectors $\dot{\mathbf{v}}_{n+1}$, $\boldsymbol{\lambda}_{n+1}$ satisfying the equilibrium conditions

$$\mathbf{M}(q_{n+1})\dot{\mathbf{v}}_{n+1} = -\mathbf{g}(q_{n+1}, \mathbf{v}_{n+1}, t_{n+1}) - \mathbf{B}^{\top}(q_{n+1})\boldsymbol{\lambda}_{n+1}\,, \tag{37e}$$

$$\boldsymbol{\Phi}(q_{n+1}) = \mathbf{0}\,. \tag{37f}$$

The term *generalized-α method* refers to the coefficients α_m, α_f in the update formula (37d) for the acceleration like variables $\mathbf{a}_n$. These auxiliary variables $\mathbf{a}_n$ were introduced by Chung and Hulbert (1993) who studied the time integration of unconstrained linear systems in linear spaces and proposed a one-parametric set of algorithmic parameters α_m, α_f, β and γ that may be considered as a quasi-standard for this type of methods, see Sect. 3.2 below.

Method (37) is initialized with starting values $q_0 \in G$ and $\mathbf{v}_0 \in \mathbb{R}^k$ that approximate the (consistent) initial values $q(t_0)$, $\mathbf{v}(t_0)$ in (15). The starting values $\dot{\mathbf{v}}_0$, $\mathbf{a}_0$ at acceleration level are approximations of $\dot{\mathbf{v}}(t_0) \in \mathbb{R}^k$, see (19). The convergence analysis in Sect. 4 below will show that the starting values need to be selected carefully to guarantee second order convergence in all solution components and to avoid spurious oscillations in the numerical solution $\boldsymbol{\lambda}_n$.

In practical applications, variable step size implementations with error control are expected to be superior to methods with fixed time step size h. For constrained systems in linear configuration spaces, a step size control algorithm for generalized-α methods with $\alpha_m = 0$ (*HHT-methods*, see Hilber et al. 1977) was developed in

(Géradin and Cardona 2001, Chap. 11). For this problem class, Jay and Negrut (2007) proposed a linear update formula for the auxiliary variables $\mathbf{a}_n$ to compensate a first-order error term resulting from a step size change at $t = t_n$.

An alternative approach is based on the elimination of these variables $\mathbf{a}_n$ in the multi-step representation of generalized-α methods according to Erlicher et al. (2002). Here, the algorithmic parameters α_m, α_f, β and γ have to be updated in each time step considering the step size ratio h_{n+1}/h_n, see (Brüls and Arnold 2008).

There is no straightforward extension of the results of Erlicher et al. (2002) from linear configuration spaces to the Lie group setting of the present paper. Furthermore, the analysis of the error propagation in time integration is simplified substantially if the time step size h is *fixed* for all time steps. For both reasons, the convergence analysis for generalized-α Lie group integrators (37) with *variable* time step size h_n will be a topic of future research that is beyond the scope of the present paper.

3.2 The Generalized-α Method in Linear Spaces

For linear configuration spaces $G = \mathbb{R}^k$ and unconstrained systems (15) with constant mass matrix $\mathbf{M}$, the generalized-α Lie group method (37) coincides with the "classical" generalized-α method that goes back to the work of Chung and Hulbert (1993). Multiplying (37d) by the (constant) mass matrix $\mathbf{M}$ and eliminating vectors $\mathbf{\Delta q}_n$ and $\dot{\mathbf{v}}_{n+1}$, we get

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h\mathbf{v}_n + (0.5 - \beta)h^2\mathbf{a}_n + \beta h^2\mathbf{a}_{n+1}\,, \tag{38a}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + (1-\gamma)h\mathbf{a}_n + \gamma h\mathbf{a}_{n+1}\,, \tag{38b}$$

$$\mathbf{0} = (1-\alpha_m)\mathbf{M}\mathbf{a}_{n+1} + \alpha_m\mathbf{M}\mathbf{a}_n + (1-\alpha_f)\mathbf{g}_{n+1} + \alpha_f\mathbf{g}_n \tag{38c}$$

with $\mathbf{g}_n := \mathbf{g}(\mathbf{q}_n, \mathbf{v}_n, t_n)$ and vectors $\mathbf{q}_n, \mathbf{q}_{n+1} \in \mathbb{R}^k$ that are typeset in boldface font to indicate the *linear* structure of the configuration space.

For a local error analysis, we suppose that $\mathbf{a}_n$ approximates $\dot{\mathbf{v}}(t_n + \Delta_\alpha h)$ with a fixed offset $\Delta_\alpha \in \mathbb{R}$, see (Jay and Negrut 2008, Sect. 2), and substitute in (38) the numerical solution vectors $\mathbf{q}_n$, $\mathbf{v}_n$, $\mathbf{a}_n$, $\mathbf{g}_n$ by $\mathbf{q}(t_n)$, $\mathbf{v}(t_n)$, $\dot{\mathbf{v}}(t_n + \Delta_\alpha h)$ and $-\mathbf{M}\dot{\mathbf{v}}(t_n)$, respectively. The resulting residuals define local truncation errors $\mathbf{l}_n^{\mathbf{q}}$, $\mathbf{l}_n^{\mathbf{v}}$ and $\mathbf{l}_n^{\mathbf{a}}$:

$$\begin{aligned}\mathbf{q}(t_{n+1}) = \mathbf{q}(t_n) + h\mathbf{v}(t_n) + (0.5-\beta)h^2\dot{\mathbf{v}}(t_n + \Delta_\alpha h) +& \\ + \beta h^2\dot{\mathbf{v}}(t_{n+1} + \Delta_\alpha h) + \mathbf{l}_n^{\mathbf{q}}\,,&\end{aligned} \tag{39a}$$

$$\mathbf{v}(t_{n+1}) = \mathbf{v}(t_n) + (1-\gamma)h\dot{\mathbf{v}}(t_n + \Delta_\alpha h) + \gamma h\dot{\mathbf{v}}(t_{n+1} + \Delta_\alpha h) + \mathbf{l}_n^{\mathbf{v}}\,, \tag{39b}$$

$$\begin{aligned}\mathbf{M}\mathbf{l}_n^{\mathbf{a}} = (1-\alpha_m)\mathbf{M}\dot{\mathbf{v}}(t_{n+1} + \Delta_\alpha h) + \alpha_m\mathbf{M}\dot{\mathbf{v}}(t_n + \Delta_\alpha h) -& \\ - (1-\alpha_f)\mathbf{M}\dot{\mathbf{v}}(t_{n+1}) - \alpha_f\mathbf{M}\dot{\mathbf{v}}(t_n)\,.&\end{aligned} \tag{39c}$$

For sufficiently smooth solutions $\mathbf{q}(t)$, the local truncation errors in (39) may be analysed by Taylor expansion of functions $\mathbf{q}(t)$, $\mathbf{v}(t)$ and $\dot{\mathbf{v}}(t)$ at $t = t_n$:

$$\mathbf{l}_n^{\mathbf{q}} = C_q h^3 \dddot{\mathbf{v}}(t_n) + \mathcal{O}(h^4) \;\text{ with }\; C_q := (1 - 6\beta - 3\Delta_\alpha)/6\,, \tag{40a}$$

$$\mathbf{l}_n^{\mathbf{v}} = (0.5 - \Delta_\alpha - \gamma)h^2 \ddot{\mathbf{v}}(t_n) + \mathcal{O}(h^3)\,, \tag{40b}$$

$$\mathbf{l}_n^{\mathbf{a}} = \big(\Delta_\alpha - (\alpha_m - \alpha_f)\big)h\ddot{\mathbf{v}}(t_n) + \mathcal{O}(h^2)\,. \tag{40c}$$

We get local truncation errors $\mathbf{l}_n^{\mathbf{q}} = \mathcal{O}(h^3)$, $\mathbf{l}_n^{\mathbf{v}} = \mathcal{O}(h^3)$ and $\mathbf{l}_n^{\mathbf{a}} = \mathcal{O}(h^2)$ if the algorithmic parameters satisfy the order condition

$$\gamma = 0.5 - \Delta_\alpha \;\text{ with }\; \Delta_\alpha := \alpha_m - \alpha_f\,. \tag{41}$$

Chung and Hulbert (1993) studied the scalar test equation $\ddot{q} + \omega^2 q = 0$ with periodic analytical solutions $q(t) = c_1 \sin\omega t + c_2 \cos\omega t$ and observed that (38) results in a frequency-dependent linear mapping $(q_n, v_n, a_n) \mapsto (q_{n+1}, v_{n+1}, a_{n+1})$. Scaling the update formulae (38a, 38b) by factors $1/h^2$ and $1/h$, respectively, we get

$$\underbrace{\begin{pmatrix} \frac{1}{(h\omega)^2} & 0 & -\beta \\ 0 & 1 & -\gamma \\ 1-\alpha_f & 0 & 1-\alpha_m \end{pmatrix}}_{=:\,\mathbf{T}_{h\omega}^+} \underbrace{\begin{pmatrix} \omega^2 q_{n+1} \\ \frac{1}{h} v_{n+1} \\ a_{n+1} \end{pmatrix}}_{=\,\mathbf{z}_{n+1}} = \underbrace{\begin{pmatrix} \frac{1}{(h\omega)^2} & 1 & 0.5-\beta \\ 0 & 1 & 1-\gamma \\ -\alpha_f & 0 & -\alpha_m \end{pmatrix}}_{=:\,\mathbf{T}_{h\omega}^0} \underbrace{\begin{pmatrix} \omega^2 q_n \\ \frac{1}{h} v_n \\ a_n \end{pmatrix}}_{=:\,\mathbf{z}_n}.$$

Recursive application yields $\mathbf{z}_n = \mathbf{T}_{h\omega}^n \mathbf{z}_0$ with $\mathbf{T}_{h\omega} := (\mathbf{T}_{h\omega}^+)^{-1}\mathbf{T}_{h\omega}^0$. Therefore, the stability and (numerical) damping properties of the generalized-α method (38) applied to $\ddot{q} + \omega^2 q = 0$ may be characterized by an eigenvalue analysis of $\mathbf{T}_{h\omega} \in \mathbb{R}^{3\times3}$. Chung and Hulbert (1993) propose to choose a user-defined parameter $\rho_\infty \in [0, 1]$ to characterize the numerical damping properties in the limit case $h\omega \to \infty$. They show that the algorithmic parameters α_m, α_f, β and γ may be defined such that the order condition (41) is satisfied and the spectral radius $\varrho(\mathbf{T}_{h\omega})$ is monotonically decreasing for $h\omega \in (0, +\infty)$ with $\lim_{h\omega\to0} \varrho(\mathbf{T}_{h\omega}) = 1$ and $\varrho(\mathbf{T}_\infty) = \rho_\infty$:

$$\alpha_m = \frac{2\rho_\infty - 1}{\rho_\infty + 1}\,,\;\; \alpha_f = \frac{\rho_\infty}{\rho_\infty + 1}\,,\;\; \gamma = \frac{1}{2} + \alpha_f - \alpha_m\,,\;\; \beta = \frac{1}{4}(\gamma + \frac{1}{2})^2\,. \tag{42}$$

For these parameters, all three eigenvalues of $\mathbf{T}_{h\omega} = \mathbf{T}_{h\omega}(\rho_\infty)$ coincide in the limit case $h\omega \to \infty$ and the Jordan canonical form of $\mathbf{T}_\infty(\rho_\infty) \in \mathbb{R}^{3\times3}$ consists of a single 3×3 Jordan block for the eigenvalue $\mu := -\rho_\infty$, i.e., $\mathbf{T}_\infty(\rho_\infty) = \mathbf{X}(\mu)\mathbf{J}(\mu)\mathbf{X}^{-1}(\mu)$ with

$$\mathbf{J}(\mu) := \begin{pmatrix} \mu & 1 & 0 \\ 0 & \mu & 1 \\ 0 & 0 & \mu \end{pmatrix},\quad \mathbf{X}(\mu) := \begin{pmatrix} 1-\mu^2 & -(2+\mu) & 0 \\ 0 & \frac{1}{2}\frac{1+\mu}{1-\mu} & -\frac{1}{(1-\mu)^2} \\ 0 & 1 & 0 \end{pmatrix}.$$

With algorithmic parameters α_m, α_f, β and γ according to (42) and a damping parameter $\rho_\infty < 1$, the linear stability of the generalized-α method (38) is always guaranteed. For the test equation $\ddot{q} + \omega^2 q = 0$, the numerical solution $(q_n, v_n, a_n)^\top$

will finally be damped out for any starting values q_0, v_0, a_0 since $\mathbf{z}_n = \mathbf{T}_{h\omega}^n(\rho_\infty)\mathbf{z}_0$ and $\lim_{n\to\infty} \mathbf{T}_{h\omega}^n(\rho_\infty) = \mathbf{0}$ because $\varrho(\mathbf{T}_{h\omega}(\rho_\infty)) < 1$, ($h\omega \in (0,\infty)$).

In a transient phase, however, $\|\mathbf{z}_n\|$ may be much larger than $\|\mathbf{z}_0\|$ since $\|\mathbf{T}^n\|$ may be much larger than $(\varrho(\mathbf{T}))^n$ for matrices that are not diagonalisable (*non-normal* matrices). Typical values are $\max_n \|\mathbf{T}^n\|_2 = \|\mathbf{T}^3\|_2 = 7.4$ for $\mathbf{T} = \mathbf{T}_\infty(\rho_\infty)$ with $\rho_\infty = 0.6$ and $\max_n \|\mathbf{T}^n\|_2 = \|\mathbf{T}^{14}\|_2 = 34.3$ for $\mathbf{T} = \mathbf{T}_\infty(\rho_\infty)$ with $\rho_\infty = 0.9$. In structural dynamics, this phenomenon is called *overshooting* since $|q_n|$ may grow rapidly in a transient phase before the numerical dissipation results finally in $\lim_{n\to 0} q_n = 0$. Overshooting is a well-known problem of unconditionally stable Newmark-type methods with second-order accuracy (Hilber and Hughes 1978) and may be a motivation to prefer first-order accurate Newmark integrators in industrial multibody system simulation (Sanborn et al. 2014).

In the quantitative error analysis, we denote the global errors of the generalized-α method in linear spaces by $\mathbf{e}_n^{(\bullet)}$ with $(\bullet)(t_n) = (\bullet)_n + \mathbf{e}_n^{(\bullet)}$. For the auxiliary vectors $\mathbf{a}_n$ that do not have a corresponding component of the analytical solution, we take into account the offset parameter Δ_α from (41) and define the global error $\mathbf{e}_n^{\mathbf{a}}$ by $\dot{\mathbf{v}}(t_n + \Delta_\alpha h) = \mathbf{a}_n + \mathbf{e}_n^{\mathbf{a}}$. For the scalar test equation $\ddot{q} + \omega^2 q = 0$, these global errors as well as the local errors l_n^q, l_n^v, l_n^a are scalar quantities and $\mathbf{T}_{h\omega}^+ \mathbf{z}_{n+1} = \mathbf{T}_{h\omega}^0 \mathbf{z}_n$ implies

$$\mathbf{T}_{h\omega}^+ \begin{pmatrix} \omega^2 e_{n+1}^q \\ \dfrac{1}{h} e_{n+1}^v \\ e_{n+1}^a \end{pmatrix} = \mathbf{T}_{h\omega}^0 \begin{pmatrix} \omega^2 e_n^q \\ \dfrac{1}{h} e_n^v \\ e_n^a \end{pmatrix} + \begin{pmatrix} \dfrac{1}{h^2} l_n^q \\ \dfrac{1}{h} l_n^v \\ l_n^a \end{pmatrix}, \tag{43}$$

see (39). As before, the first and second row are scaled by $1/h^2$ and $1/h$, respectively. The resulting first-order error term $l_n^q/h^2 = C_q h \ddot{v}(t_n) + \mathcal{O}(h^2)$ may strongly affect the result accuracy.

This order reduction phenomenon is known from the convergence analysis for the application of Newmark-type methods to constrained mechanical systems in linear configuration spaces, see (Cardona and Géradin 1994). In the limit case $\omega \to \infty$, the transient solution behaviour is dominated by an oscillating first-order error term that is finally damped out by numerical dissipation. To study this qualitative solution behaviour in full detail, we introduce a new variable $\lambda := \omega^2 q$ and rewrite the test equation as a singular singularly perturbed problem with perturbation parameter $\varepsilon := 1/\omega$, see (Lubich 1993):

$$\ddot{q} + \omega^2 q = 0 \quad \Leftrightarrow \quad \left.\begin{aligned} \ddot{q} &= -\lambda \\ \frac{1}{\omega^2}\lambda &= q \end{aligned}\right\} \tag{44}$$

The corresponding reduced system ($\varepsilon = 0$, i.e., $\omega \to \infty$) is a constrained system (15) with $G = \mathbb{R}$ and $k = m = 1$:

$$\left.\begin{aligned} \ddot{q} &= -\lambda \\ 0 &= q \end{aligned}\right\} \tag{45}$$

With the notation $\lambda_n := \omega^2 q_n$, the generalized-$\alpha$ method (38) for the singularly perturbed system (44) converges for $\omega \to \infty$ to the generalized-α method (37) for the constrained system (45) and we get in (43) both for finite frequencies ω and in the limit case $\omega \to \infty$:

$$\mathbf{T}^{+}_{h\omega}\mathbf{e}^{\mathbf{r}}_{n+1} = \mathbf{T}^{0}_{h\omega}\mathbf{e}^{\mathbf{r}}_{n} + \mathbf{l}^{\mathbf{r}}_{n} \tag{46}$$

with

$$\mathbf{e}^{\mathbf{r}}_{n} := \begin{pmatrix} e^{\lambda}_{n} \\ r_n \\ e^{a}_{n} \end{pmatrix}, \quad \mathbf{l}^{\mathbf{r}}_{n} := \begin{pmatrix} 0 \\ \dfrac{1}{h} l^{v}_{n} + \dfrac{l^{q}_{n+1} - l^{q}_{n}}{h^2} \\ l^{a}_{n} \end{pmatrix} \tag{47}$$

and

$$r_n := \frac{1}{h}\big(e^{v}_{n} + \frac{1}{h} l^{q}_{n}\big) = \frac{1}{h}\big(e^{v}_{n} + C_q h^2 \ddot{v}(t_n)\big) + \mathcal{O}(h^2)\,. \tag{48}$$

The error recursion in terms of e^{λ}_{n}, r_n and e^{a}_{n} provides the basis for a detailed convergence analysis:

Theorem 3.1 *Consider the time discretization of the linear test equations (44) and (45) by a generalized-α method with parameters α_m, α_f, β and γ according to (42) for some numerical damping parameter $\rho_\infty \in [0, 1)$.*

(a) The discretization errors are bounded by

$$\|\mathbf{l}^{\mathbf{r}}_{n}\| = \mathcal{O}(h^2)\,, \quad \|\mathbf{e}^{\mathbf{r}}_{n+1} - \mathbf{T}_{h\omega}\mathbf{e}^{\mathbf{r}}_{n}\| = \mathcal{O}(h^2)\,, \tag{49}$$

$$\|\mathbf{e}^{\mathbf{r}}_{n} - \mathbf{T}^{n}_{h\omega}\mathbf{e}^{\mathbf{r}}_{0}\| = \mathcal{O}(h^2) \tag{50}$$

and

$$\|\mathbf{e}^{\mathbf{r}}_{n}\| \le \|\mathbf{T}^{n}_{h\omega}\|\,\|\mathbf{e}^{\mathbf{r}}_{0}\| + \mathcal{O}(h^2)\,. \tag{51}$$

(b) For starting values $\lambda_0 = \lambda(t_0) + \mathcal{O}(h^2)$, $a_0 = \dot{v}(t_0 + \Delta_\alpha h) + \mathcal{O}(h^2)$, we have $\|\mathbf{e}^{\mathbf{r}}_{0}\| = \mathcal{O}(h)$ if $v_0 = v(t_0) + \mathcal{O}(h^2)$. This error estimate may be improved by one power of h perturbing the starting value v_0 such that

$$v_0 = v(t_0) + C_q h^2 \ddot{v}(t_0) + \mathcal{O}(h^3)\,. \tag{52}$$

In that case, we get $\|\mathbf{e}^{\mathbf{r}}_{n}\| = \mathcal{O}(h^2)$, ($n \ge 0$).

Proof (a) Because of

$$l^{q}_{n+1} - l^{q}_{n} = C_q h^3\big(\ddot{v}(t_{n+1}) - \ddot{v}(t_n)\big) + \mathcal{O}(h^4) = \mathcal{O}(h^4)\,,$$

the local error term $\mathbf{l}_n^{\mathbf{r}}$ is of size $\mathcal{O}(h^2)$, see (40), and (49) is a direct consequence of the error recursion (46). The assumptions on parameters $\alpha_m, \alpha_f, \beta$ and γ imply $\varrho(\mathbf{T}_{h\omega}) < 1$ and the existence of a norm $\|\mathbf{T}\|_\rho$ with $\kappa := \|\mathbf{T}_{h\omega}\|_\rho < 1$, see, e.g., (Quarteroni et al. 2000, Sect. 1.11.1). Therefore,

$$\begin{aligned}\|\mathbf{e}_n^{\mathbf{r}} - \mathbf{T}_{h\omega}^n \mathbf{e}_0^{\mathbf{r}}\|_\rho &\le \|\mathbf{e}_n^{\mathbf{r}} - \mathbf{T}_{h\omega}\mathbf{e}_{n-1}^{\mathbf{r}}\|_\rho + \|\mathbf{T}_{h\omega}\mathbf{e}_{n-1}^{\mathbf{r}} - \mathbf{T}_{h\omega}^n \mathbf{e}_0^{\mathbf{r}}\|_\rho \\ &\le \|\mathbf{e}_n^{\mathbf{r}} - \mathbf{T}_{h\omega}\mathbf{e}_{n-1}^{\mathbf{r}}\|_\rho + \|\mathbf{T}_{h\omega}\|_\rho \, \|\mathbf{e}_{n-1}^{\mathbf{r}} - \mathbf{T}_{h\omega}^{n-1} \mathbf{e}_0^{\mathbf{r}}\|_\rho \\ &\le Ch^2 + \kappa \|\mathbf{e}_{n-1}^{\mathbf{r}} - \mathbf{T}_{h\omega}^{n-1} \mathbf{e}_0^{\mathbf{r}}\|_\rho\end{aligned}$$

with an appropriate constant $C > 0$, see (49). Recursive application of this error estimate results in

$$\|\mathbf{e}_n^{\mathbf{r}} - \mathbf{T}_{h\omega}^n \mathbf{e}_0^{\mathbf{r}}\|_\rho \le \sum_{i=0}^{n-1} \kappa^i \, Ch^2 + \kappa^n \|\mathbf{e}_0^{\mathbf{r}} - \mathbf{T}_{h\omega}^0 \mathbf{e}_0^{\mathbf{r}}\|_\rho < \frac{1}{1-\kappa} \, Ch^2$$

and (50) follows from the equivalence of all norms in the finite dimensional space $\mathbb{R}^3$. Error bound (51) is a straightforward consequence of the triangle inequality.

(b) We get $\|\mathbf{e}_0^{\mathbf{r}}\| = |r_0| + \mathcal{O}(h^2)$ and the estimates for $\|\mathbf{e}_0^{\mathbf{r}}\|$ and for $\|\mathbf{e}_n^{\mathbf{r}}\|$, ($n > 0$), follow from the definition of r_n, see (48), and from part (a) of the theorem. □

The most natural choice of starting values $\lambda_0 := \lambda(t_0)$, $v_0 := v(t_0)$, $a_0 := \dot{v}(t_0 + \Delta_\alpha h)$ yields $\mathbf{e}_0^{\mathbf{r}} = (\, 0, \, r_0, \, 0\,)^\top$ with $r_0 = C_q h \ddot{v}(t_0) + \mathcal{O}(h^2)$, see (48). In error estimate (50), we obtain for $\ddot{v}(t_0) \neq 0$ a first-order error term being amplified by matrix-valued factors $\mathbf{T}_{h\omega}^n$ that are well known from the analysis of the "overshoot" phenomenon by Hilber and Hughes (1978). In the limit case $h\omega \to \infty$, this term may be studied in more detail using the Jordan canonical form of $\mathbf{T}_\infty$, see (Cardona and Géradin 1989, 1994). We get

$$\mathbf{T}_\infty^n \, \mathbf{e}_0^{\mathbf{r}} = \mathbf{X}(-\rho_\infty) \, \mathbf{J}^n(-\rho_\infty) \, \mathbf{X}^{-1}(-\rho_\infty) \, \mathbf{e}_0^{\mathbf{r}}$$

with the Jordan block $\mathbf{J}(-\rho_\infty) \in \mathbb{R}^{3\times 3}$. It may be verified by induction that the non-zero elements of $\mathbf{J}^n(-\rho_\infty)$ are given by $(-\rho_\infty)^n$, $n(-\rho_\infty)^{n-1}$ and $n(n-1)(-\rho_\infty)^{n-2}/2$. Straightforward computations show that the global error e_n^λ (that coincides up to a term of size $\mathcal{O}(h^2)$ with the first component of $\mathbf{T}_\infty^n \mathbf{e}_0^{\mathbf{r}}$) satisfies $e_n^\lambda = c_n h \ddot{v}(t_0) + \mathcal{O}(h^2)$ with

$$c_n := C_q (1+\rho_\infty)^2 \Big(\frac{n}{2}(n-1)(\rho_\infty^2 - 1)(-\rho_\infty)^{n-2} + n(2-\rho_\infty)(-\rho_\infty)^{n-1} \Big) \, . \tag{53}$$

After a transient phase, the first-order error term $c_n h \ddot{v}(t_0)$ is damped out since $\lim_{n\to\infty} c_n = 0$ for any $\rho_\infty \in [0, 1)$. In the transient phase, however, the error constants c_n may become very large with maximum absolute values of size $|c_3| = 6.8$ for $\rho_\infty = 0.6$, $|c_{15}| = 31.9$ for $\rho_\infty = 0.9$ and $|c_{161}| = 334.3$ for $\rho_\infty = 0.99$.

For the test equation (45) itself, this error analysis has not much practical relevance since $q(t) \equiv 0$ implies $\ddot{v}(t) \equiv 0$ and $\mathbf{e}_0^{\mathbf{r}} = \mathbf{0}$ for exact starting values $\lambda_0 = \lambda(t_0) = 0$,

$v_0 = v(t_0) = 0$, $a_0 = \dot{v}(t_0 + \Delta_\alpha h) = 0$. Substituting the trivial constraint $q = 0$ by a rheonomic constraint $q(t) = t^3/6$, we may construct, however, a slightly more complex test problem with non-vanishing first-order error term $r_0 = C_q h$ since $l_n^q = C_q h^3 \ddot{v}(t_n) = C_q h^3$ and the local truncation errors l_n^v, l_n^a vanish identically. For this test problem, the global error in λ really suffers from order reduction since $e_n^\lambda = c_n h$.

The convergence analysis for generalized-α methods shows that this order reduction phenomenon is typical for the initialization of method (37) with exact starting values $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}(t_0)$, $\mathbf{v}_0 = \mathbf{v}(t_0)$ and $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0 + \Delta_\alpha h)$, see (Arnold et al. 2015) and Sect. 4 below. For linear configuration spaces ($G = \mathbb{R}^k$), the global error in $\boldsymbol{\lambda}$ is bounded by

$$[\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^\top]\big(\mathbf{q}(t_n)\big)\,\mathbf{e}_n^{\boldsymbol{\lambda}} = c_n h\,\mathbf{B}\big(\mathbf{q}(t_0)\big)\ddot{\mathbf{v}}(t_0) + \mathcal{O}(h^2) \tag{54}$$

with the error constants c_n being defined in (53). The undesired first-order error term is nicely illustrated by numerical test results for the mathematical pendulum, see (Arnold et al. 2015, Sect. 2.3):

Example 3.2 Consider a mathematical pendulum of mass m and length l in Cartesian coordinates $\mathbf{q} = (x, y)^\top$ with constraint $(x^2 + y^2 - l^2)/2 = 0$, see (15c). In (15), we have $\mathbf{M} = m\mathbf{I}_2$, $\mathbf{g} = (\,0\,,\,g\,)^\top$ with $m = l = 1$, $g = 9.81$ (here and in the following, all physical units are omitted). We fix the total energy $E = m(\dot{x}_0^2 + \dot{y}_0^2)/2 + mgy_0$ to $E = m/2 - mgl$ and determine the consistent initial values x_0, y_0, $\dot{x}_0$, $\dot{y}_0$ and λ_0 by the initial deviation x_0 from the equilibrium position.

Method (37) is applied with algorithmic parameters according to (42) and damping parameter $\rho_\infty = 0.9$. The starting values are set to $\mathbf{q}_0 := (x_0, y_0)^\top$, $\mathbf{v}_0 := (\dot{x}_0, \dot{y}_0)^\top$ and $\dot{\mathbf{v}}_0 := (\ddot{x}_0, \ddot{y}_0)^\top$ with accelerations $\ddot{x}_0$, $\ddot{y}_0$ that are obtained from evaluating the equations of motion for the consistent initial values x_0, y_0, $\dot{x}_0$, $\dot{y}_0$, λ_0. The acceleration like variables $\mathbf{a}_n$ are initialized with $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0) + \Delta_\alpha h \ddot{\mathbf{v}}(t_0) + \mathcal{O}(h^2) = \dot{\mathbf{v}}(t_0 + \Delta_\alpha h) + \mathcal{O}(h^2)$ using the starting value $\dot{\mathbf{v}}_0 = \dot{\mathbf{v}}(t_0)$ and a difference approximation of $\ddot{\mathbf{v}}(t_0)$.

Figure 5 shows on a short time interval the global error in λ for initial values $x_0 = 0$ (marked by dots) and $x_0 = 0.2$ (marked by "+") for two different step sizes h. If we start in the equilibrium position, the error is very small but for $x_0 = 0.2$, the oscillating

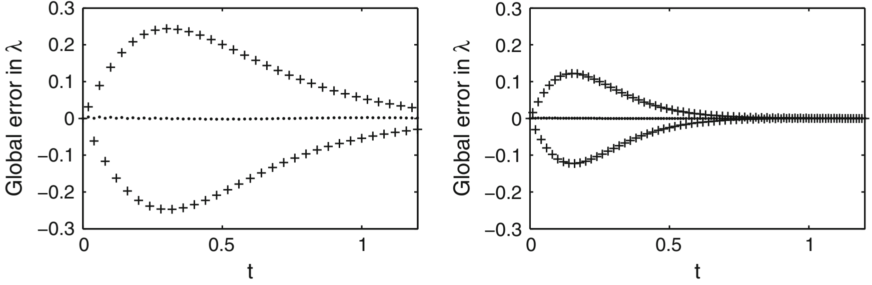


Fig. 5 Mathematical pendulum: Global error in λ for $x_0 = 0$ ("·") and $x_0 = 0.2$ ("+")

error in λ reaches a maximum amplitude of 2.48×10^{-1} for $h = 2.0 \times 10^{-2}$ and 1.23×10^{-1} for $h = 1.0 \times 10^{-2}$. After about 100 time steps these transient errors are damped out.

The numerical results in Fig. 5 show that in the transient phase the generalized-α method (37) may suffer from spurious oscillations of amplitude $\mathcal{O}(h)$. According to (54), this first-order error term is given by $c_n h\, \mathbf{B}\big(\mathbf{q}(t_0)\big)\ddot{\mathbf{v}}(t_0)$ with $\mathbf{B}\big(\mathbf{q}(t_0)\big)\ddot{\mathbf{v}}(t_0) = -3gx_0\dot{x}_0/y_0$. Therefore, the spurious oscillations and the order reduction disappear if we start at the equilibrium position $x_0 = 0$. Reducing the damping parameter ρ_∞ in (42), the oscillations are damped out more rapidly but may still be observed.

3.3 Numerical Tests for the Heavy Top Benchmark Problem

In the present section, we study the convergence behaviour of the generalized-α Lie group integrator (37) numerically. We use algorithmic parameters according to (42) with the numerical damping parameter $\rho_\infty = 0.9$ and apply (37) to the equations of motion (21), (22) of the heavy top benchmark problem in configuration spaces $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and $G = \mathrm{SE}(3)$, respectively. Initial values $q(t_0)$, $\mathbf{v}(t_0)$ are given in Sect. 2.4. In the numerical tests, the integrator was initialized with starting values $q_0 := q(t_0)$, $\mathbf{v}_0 := \mathbf{v}(t_0)$, $\dot{\mathbf{v}}_0 := \dot{\mathbf{v}}(t_0)$ and $\mathbf{a}_0 := \dot{\mathbf{v}}(t_0)$ with $\dot{\mathbf{v}}(t_0)$ denoting the consistent acceleration vector being defined in (19).

In Fig. 6, the asymptotic behaviour of the global errors in q_n, $\mathbf{v}_n$ and $\boldsymbol{\lambda}_n$ for $h \to 0$ is visualized in terms of the maximum $\max_n \|\mathbf{e}_n^{(\bullet)}\|/\|(\bullet)_n\|$ of the norm of relative errors in the time interval $[t_0, t_{\mathrm{end}}] = [0, 1]$. Here, the numerical solutions for $h = 1.25 \times 10^{-4}$, $h = 2.5 \times 10^{-4}$, $h = 5.0 \times 10^{-4}, \ldots, h = 4.0 \times 10^{-3}$ are compared to a reference solution that has been obtained numerically with the very small time step size $h = 2.5 \times 10^{-5}$. In double logarithmic scale, the plots of global errors in q_n and $\mathbf{v}_n$ are straight lines of slope +2 (for both configuration spaces). These numerical test results indicate second-order convergence for components q and $\mathbf{v}$.

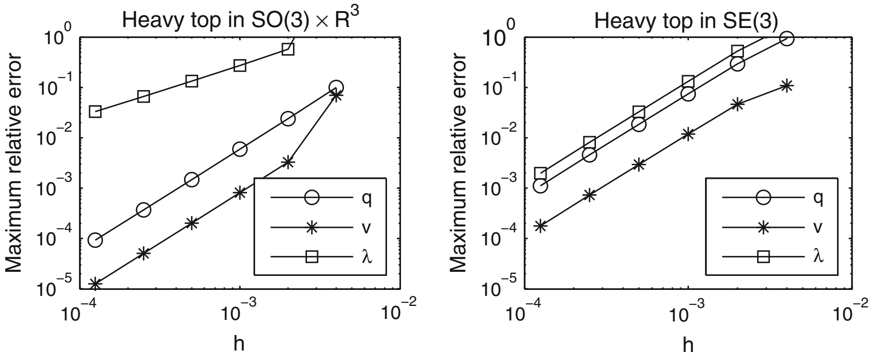


Fig. 6 Heavy top benchmark (index-3 formulation): Global error of integrator (37) versus h for $t \in [0, 1]$. *Left plot* $\mathrm{SO}(3) \times \mathbb{R}^3$, *right plot* SE(3)

The error constants depend on model parameters, initial values and configuration space. With the test setup of Sect. 2.4, the velocity components $\mathbf{v}(t)$ vary much more rapidly for $G = \mathrm{SE}(3)$ than for $G = \mathrm{SO}(3) \times \mathbb{R}^3$, see Fig. 4. This might explain the substantially larger error constants for q_n and $\mathbf{v}_n$ in the right plot of Fig. 6. For other setups, much smaller error constants have been observed for the configuration space SE(3), see, e.g., the numerical test results of Brüls et al. (2011) for a slowly rotating top with an initial angular velocity $\boldsymbol{\Omega}(0)$ that has been reduced by a factor of 100.

Note, that Fig. 6 shows the norm of *relative* errors. The rather large nominal values of $\mathbf{v}(t)$ with $\|\boldsymbol{\Omega}(0)\| \approx 150.0$ result systematically in relative errors that have a substantially smaller norm than the ones in the position coordinates $q(t)$.

For the Lagrange multipliers $\boldsymbol{\lambda}(t)$, we observe order reduction since slope $+1$ of the curve for the global errors in $\boldsymbol{\lambda}_n$ in the left plot of Fig. 6 indicates first-order convergence. The test results for $G = \mathrm{SE}(3)$ in the right plot of Fig. 6 are qualitatively different from the ones in the left plot since they indicate second-order convergence for *all* solution components. A formal proof of this numerically observed convergence behaviour will be given in Theorem 4.18 and Example 4.19 below.

Guided by the test results for the mathematical pendulum in Example 3.2, we expect that the order reduction phenomenon might affect the numerical solution only in a transient phase and the first-order error terms in $\boldsymbol{\lambda}_n$ are finally damped out by numerical dissipation. This is nicely illustrated by Fig. 7 that shows the numerical solution $\lambda_{n,1}$ for $t \in [0, 0.1]$ and two different time step sizes. In the configuration space $G = \mathrm{SO}(3) \times \mathbb{R}^3$ (solid lines), spurious oscillations are observed that are damped out after about 50 time steps and have a maximum amplitude that depends linearly on h. Beyond this transient phase, the results coincide up to plot accuracy with the dashed lines showing simulation results for the configuration space $G = \mathrm{SE}(3)$ that do not suffer from order reduction.

Neglecting the transient behaviour, we observe for both Lie group formulations second-order convergence in all solution components, see Fig. 8 that shows the maximum of the norm of global errors in time interval [0.5, 1], i.e., beyond the transient phase.

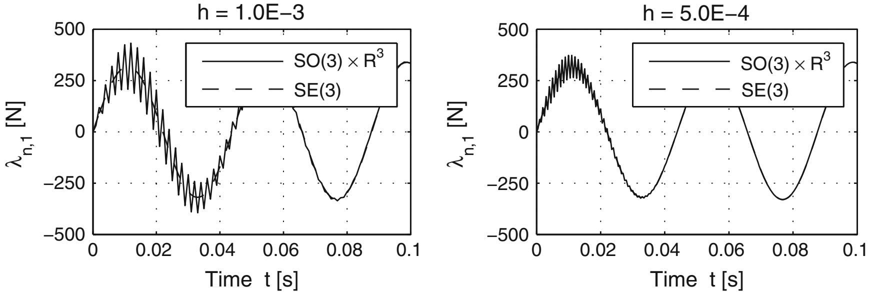


Fig. 7 Heavy top benchmark (index-3 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and $G = \mathrm{SE}(3)$): Numerical solution of Lagrange multiplier $\lambda_{n,1}$. *Left plot* $h = 1.0 \times 10^{-3}$, *right plot* $h = 5.0 \times 10^{-4}$

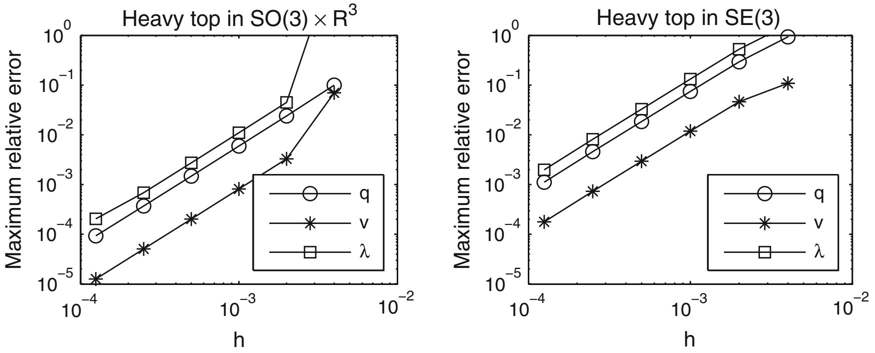


Fig. 8 Heavy top benchmark (index-3 formulation): Global error of integrator (37) versus h for $t \in [0.5, 1]$. *Left plot* $\mathrm{SO}(3) \times \mathbb{R}^3$, *right plot* $\mathrm{SE}(3)$

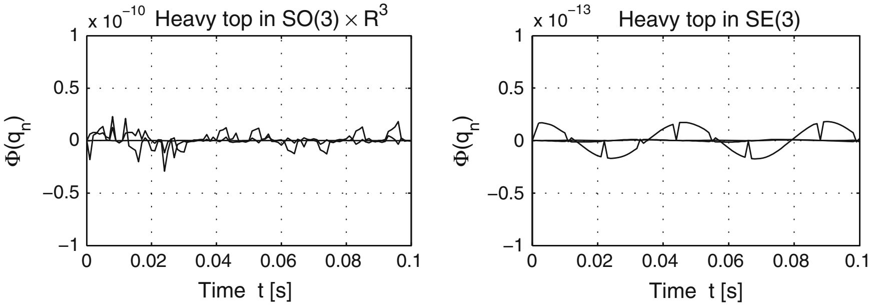


Fig. 9 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, index-3 formulation): Residuals in constraints (15c). *Left plot* $\mathrm{SO}(3) \times \mathbb{R}^3$, *right plot* $\mathrm{SE}(3)$

By construction, the Lie group integrator (37) defines a numerical solution q_n that satisfies the holonomic constraints $\mathbf{\Phi}(q) = \mathbf{0}$. In a practical implementation, the residuals remain in the size of the stopping bounds for the Newton method that is used to solve in each time step the system of nonlinear equations (37). For the numerical tests we applied a combined absolute and relative error criterion with tolerances $\mathrm{ATOL} = 10^{-10}$ for the absolute errors and $\mathrm{RTOL} = 10^{-8}$ for the relative errors and observe constraint residuals of size $\|\mathbf{\Phi}(q_n)\| \ll 10^{-10}$, see Fig. 9.

Situation is different for the residuals in the hidden constraints (16) that are in general of the size of global discretization errors since $\mathbf{B}\big(q(t)\big)\mathbf{v}(t) = \mathbf{0}$. The left plot of Fig. 10 shows these non-vanishing residuals $\mathbf{B}(q_n)\mathbf{v}_n$ for $h = 1.0 \times 10^{-3}$ and $G = \mathrm{SO}(3) \times \mathbb{R}^3$. They are of size $\|\mathbf{B}(q_n)\mathbf{v}_n\| \le 0.025$ and suffer from the transient spurious oscillations being known from Fig. 7 above. For the configuration space $G = \mathrm{SE}(3)$, the constraint residuals are smaller by eight orders of magnitude with $\max_n \|\mathbf{B}(q_n)\mathbf{v}_n\| \approx 1.0 \times 10^{-10}$. This unexpected solution behaviour is visualized in the right plot of Fig. 10. It is closely related to the fact that the constraint Jacobian $\mathbf{B}(q)$ in (22) is constant along the analytical solution $q(t)$, see Sect. 3.6 below for a more detailed analysis.

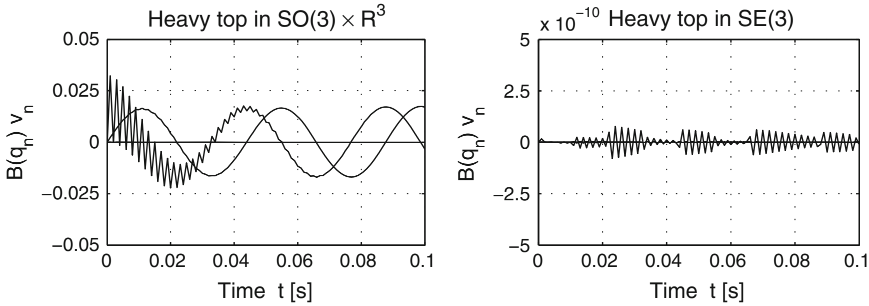


Fig. 10 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, index-3 formulation): Residuals in hidden constraints (16). *Left plot* $SO(3) \times \mathbb{R}^3$, *right plot* $SE(3)$

In all numerical tests of the present section, the numerical damping parameter was set to $\rho_\infty := 0.9$. The qualitative behaviour of the numerical solution in configuration spaces $SO(3) \times \mathbb{R}^3$ and $SE(3)$ is, however, not sensitive w.r.t. this algorithmic parameter, see, e.g., the results for $\rho_\infty = 0.8$ and the test setup of Fig. 7 in (Brüls et al. 2011) and the results for $\rho_\infty = 0.6$ and the test setup of Fig. 20 below in (Arnold et al. 2015).

3.4 Lie Group Time Integration and Index Reduction

The large amplitude of spurious oscillations in the numerical solution $\boldsymbol{\lambda}_n$, see Fig. 7, results from order reduction in Newmark-type methods that are directly applied to the index-3 formulation of the equations of motion for constrained mechanical systems, see (Cardona and Géradin 1994) and (Arnold et al. 2015). As an alternative to this direct time discretization of the index-3 Lie group DAE (15) we consider in the present section an analytical index reduction *before* time integration. We follow the approach of Gear et al. (1985) that is well known for equations of motion in linear spaces and was extended to the Lie group setting of the present paper in (Arnold et al. 2011a).

Gear et al. (1985) introduced an auxiliary vector $\boldsymbol{\eta}(t) \in \mathbb{R}^m$ in the kinematic equations to couple the hidden constraints at the level of velocity coordinates to the equations of motion. In the Lie algebra approach to Lie group time integration, these modified kinematic equations get the form $\dot{q}(t) = DL_{q(t)}(e) \cdot \widetilde{\boldsymbol{\Delta}\mathbf{q}}(t)$ with $\widetilde{\boldsymbol{\Delta}\mathbf{q}} \in \mathfrak{g}$ being defined by $\boldsymbol{\Delta}\mathbf{q} = \mathbf{v} - \mathbf{B}^\top(q)\boldsymbol{\eta}$, see (6). The resulting *stabilized index-2 formulation* of the equations of motion is given by

$$\dot{q} = DL_q(e) \cdot \widetilde{\boldsymbol{\Delta}\mathbf{q}}\,, \tag{55a}$$

$$\boldsymbol{\Delta}\mathbf{q} = \mathbf{v} - \mathbf{B}^\top(q)\boldsymbol{\eta}\,, \tag{55b}$$

$$\mathbf{M}(q)\dot{\mathbf{v}} = -\mathbf{g}(q, \mathbf{v}, t) - \mathbf{B}^\top(q)\boldsymbol{\lambda}\,, \tag{55c}$$

$$\mathbf{\Phi}(q) = \mathbf{0}\,, \tag{55d}$$

$$\mathbf{B}(q)\mathbf{v} = \mathbf{0}\,. \tag{55e}$$

For the modified kinematic equations (55a), the time derivative of the holonomic constraints (55d) is given by $\mathbf{0} = \mathbf{B}(q)\mathbf{\Delta q}$, see (16). Therefore, Eqs. (55b) and (55e) yield $\mathbf{0} = [\mathbf{B}\mathbf{B}^\top](q)\boldsymbol{\eta}$ and $\boldsymbol{\eta}(t) \equiv \mathbf{0}$ since the full rank assumption on the constraint matrix $\mathbf{B} \in \mathbb{R}^{m \times k}$ implies that $\mathbf{B}\mathbf{B}^\top \in \mathbb{R}^{m \times m}$ is non-singular. Hence, $\mathbf{\Delta q}(t) = \mathbf{v}(t)$ and the stabilized index-2 formulation (55) is analytically equivalent to the original equations of motion (15).

The index analysis of Gear et al. (1985) is extended straightforwardly from linear spaces to the Lie group setting of the present paper and shows that the analytical transformation from (15) to (55) reduces the DAE index of the equations of motion from three to two.

The generalized-α method for the index-2 system (55) satisfies at $t = t_{n+1}$ the holonomic constraints (55d) as well as the hidden constraints (55e). An auxiliary vector $\boldsymbol{\eta}_n \in \mathbb{R}^m$ is added to the definition of the increment vector $\mathbf{\Delta q}_n$, see (55b):

$$q_{n+1} = q_n \circ \exp(h\widetilde{\mathbf{\Delta q}_n})\,, \tag{56a}$$

$$\begin{aligned}\mathbf{\Delta q}_n &= \mathbf{v}_n - \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n + \\ &\quad + (0.5 - \beta)h\mathbf{a}_n + \beta h\mathbf{a}_{n+1}\,,\end{aligned} \tag{56b}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_n + (1 - \gamma)h\mathbf{a}_n + \gamma h\mathbf{a}_{n+1}\,, \tag{56c}$$

$$(1 - \alpha_m)\mathbf{a}_{n+1} + \alpha_m\mathbf{a}_n = (1 - \alpha_f)\dot{\mathbf{v}}_{n+1} + \alpha_f\dot{\mathbf{v}}_n\,, \tag{56d}$$

$$\mathbf{M}(q_{n+1})\dot{\mathbf{v}}_{n+1} = -\mathbf{g}(q_{n+1}, \mathbf{v}_{n+1}, t_{n+1}) - \mathbf{B}^\top(q_{n+1})\boldsymbol{\lambda}_{n+1}\,, \tag{56e}$$

$$\mathbf{\Phi}(q_{n+1}) = \mathbf{0}\,, \tag{56f}$$

$$\mathbf{B}(q_{n+1})\mathbf{v}_{n+1} = \mathbf{0}\,. \tag{56g}$$

Following the test scenario of Sect. 3.3, we study the asymptotic behaviour of integrator (56) for $h \to 0$ by numerical tests for the heavy top benchmark in configuration spaces $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and $G = \mathrm{SE}(3)$, respectively. As before, we scale the norm of the (absolute) global errors by the norm of nominal values and consider the maximum of these relative errors in time interval $[t_0, t_{\mathrm{end}}] = [0, 1]$. Figure 11 shows these maximum values of the norm of global errors in q_n, $\mathbf{v}_n$ and $\boldsymbol{\lambda}_n$ versus time step size h. In double logarithmic scale, we get in the step size range $h \geq 2.5 \times 10^{-4}$ curves of slope $+2$ indicating second-order error terms in all solution components.

For the configuration space $\mathrm{SO}(3) \times \mathbb{R}^3$ (left plot) and very small time step sizes $h < 2.5 \times 10^{-4}$, the errors in $\boldsymbol{\lambda}_n$ are dominated by a first-order term. On the other hand, the error constants of the second-order error terms are slightly smaller than the ones in the corresponding plots for the index-3 integrator (37), see Figs. 6 and 8. The results for configuration space $\mathrm{SE}(3)$ in the right plot of Fig. 11 coincide up to plot accuracy with the ones in Figs. 6 and 8.

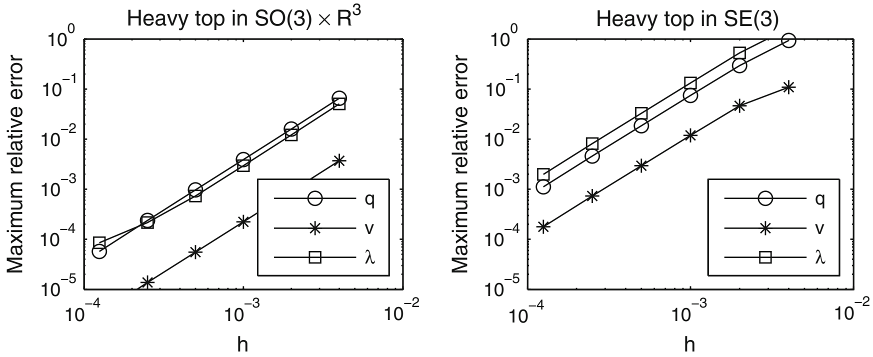


Fig. 11 Heavy top benchmark (stabilized index-2 formulation): Global error of integrator (56) versus h for $t \in [0, 1]$. *Left plot* $\mathrm{SO}(3) \times \mathbb{R}^3$, *right plot* $\mathrm{SE}(3)$

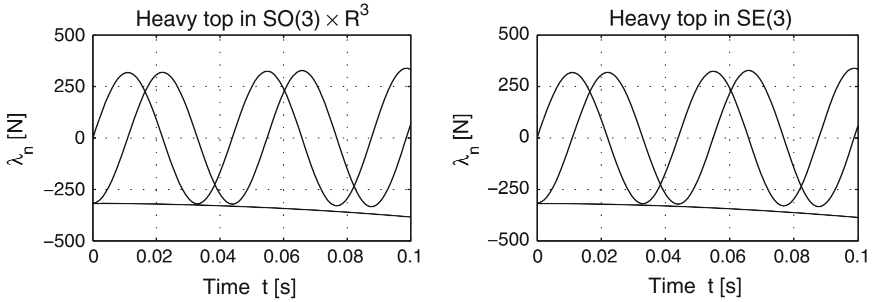


Fig. 12 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, stabilized index-2 formulation): Numerical solution $\boldsymbol{\lambda}_n$. *Left plot* $\mathrm{SO}(3) \times \mathbb{R}^3$, *right plot* $\mathrm{SE}(3)$

The comparison of time histories for $\boldsymbol{\lambda}_n$ in Figs. 7 and 12 shows that the spurious oscillations seem to disappear if hidden constraints are taken into account for time integration, see (56g). For a more detailed analysis, we consider in Fig. 13 the relative global error in $\lambda_{n,1}$ for $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and two different time step sizes. There is an oscillating first-order error term of maximum amplitude $0.64\,h$ that is rapidly damped out. For time step sizes $h \geq 5.0 \times 10^{-4}$, it does not contribute significantly to the overall global error in $\boldsymbol{\lambda}_n$ on time interval $[0, 1]$ that is approximately of size $3.0 \times 10^3\, h^2$, see Fig. 11.

The test results in the right plot of Fig. 10 indicate that the index-3 integrator (37) yields for the heavy top benchmark in $G = \mathrm{SE}(3)$ a numerical solution q_n, $\mathbf{v}_n$ that satisfies the hidden constraints (56g) up to (very) small residuals. Therefore, the auxiliary variables $\boldsymbol{\eta}_n \in \mathbb{R}^m$ that represent the differences between integrators (37) and (56) vanish in that case identically, see also Sect. 3.6 below.

For the configuration space $G = \mathrm{SO}(3) \times \mathbb{R}^3$, we observed in the left plot of Fig. 10 non-vanishing constraint residuals $\mathbf{B}(q_n)\mathbf{v}_n$ for the index-3 integrator (37). In integrator (56), they are compensated by auxiliary variables $\boldsymbol{\eta}_n = \mathcal{O}(h^2)$ for the stabilized index-2 formulation of the equations of motion. Figure 14 shows $\boldsymbol{\eta}_n$ versus

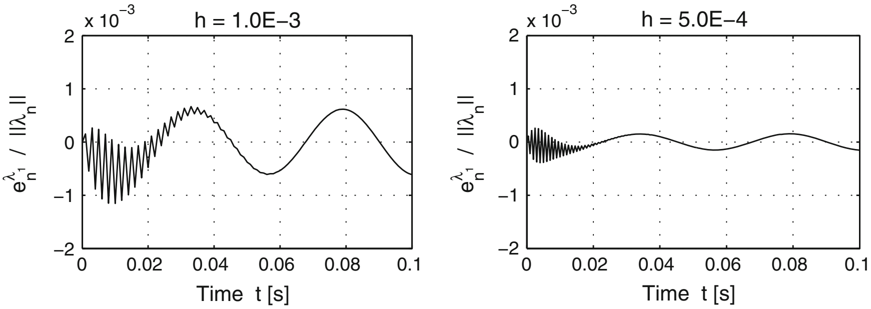


Fig. 13 Heavy top benchmark (stabilized index-2 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$): Global error $e_n^{\lambda_1}/\|\boldsymbol{\lambda}_n\|$. *Left plot* $h = 1.0 \times 10^{-3}$, *right plot* $h = 5.0 \times 10^{-4}$

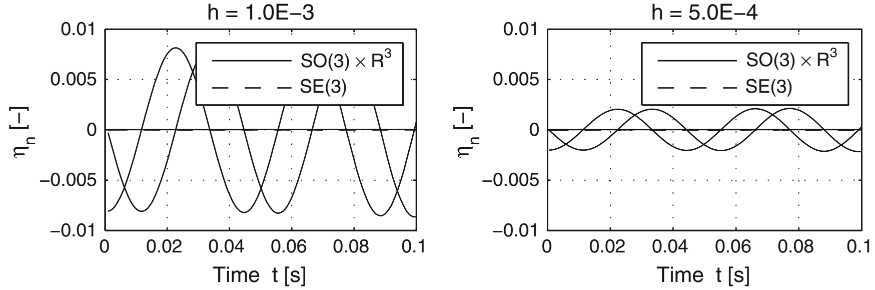


Fig. 14 Heavy top benchmark (stabilized index-2 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and $G = \mathrm{SE}(3)$): Numerical solution $\boldsymbol{\eta}_n$. *Left plot* $h = 1.0 \times 10^{-3}$, *right plot* $h = 5.0 \times 10^{-4}$

t_n for two different time step sizes. The maximum amplitudes of $\boldsymbol{\eta}_n$ differ by a factor of 4 if step sizes h and $h/2$ are considered, $h = 1.0 \times 10^{-3}$. Therefore, we expect second-order convergence for solution components $\boldsymbol{\eta}_n$.

Finally, we study the constraint residuals for a practical implementation of integrator (56). As before, the residuals in the holonomic constraints (15c) at the level of position coordinates are very small. For the hidden constraints (16) at the level of velocity coordinates, the residuals for integrator (56) are shown in Fig. 15. For the heavy top benchmark, they are of size 2.0×10^{-9} for $G = \mathrm{SO}(3) \times \mathbb{R}^3$ and of size 2.0×10^{-15} for $G = \mathrm{SE}(3)$.

In all these numerical tests for integrator (56), the extra effort for considering the hidden constraints (16) helps to reduce systematically shortcomings like spurious oscillations that were observed for the index-3 integrator (37) in Sect. 3.3.

3.5 Implementation Aspects

In each time step, the generalized-α method (37) defines the numerical solution $(q_{n+1}, \mathbf{v}_{n+1}, \dot{\mathbf{v}}_{n+1}, \mathbf{a}_{n+1}, \boldsymbol{\lambda}_{n+1})$ implicitly by a mixed system of linear and nonlinear

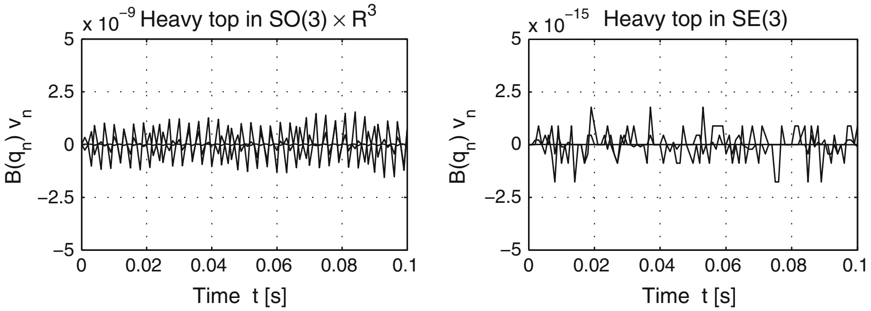


Fig. 15 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, stabilized index-2 formulation): Residuals in hidden constraints (16). *Left plot* $\mathrm{SO}(3) \times \mathbb{R}^3$, *right plot* $\mathrm{SE}(3)$

equations in $G \times \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^m$. Despite the nonlinear structure of the configuration space G, these equations may be solved numerically by a Newton–Raphson iteration in a *linear* space expressing $q_{n+1} \in G$ in terms of $\widetilde{\mathbf{\Delta q}}_n \in \mathfrak{g}$.

For the practical implementation of this Lie algebra approach, the Newton–Raphson method has to be combined with an appropriate scaling of equations and unknowns to guarantee that the condition number of the iteration matrix is bounded independently of h, see (Petzold and Lötstedt 1986) and the more recent discussion in (Bottasso et al. 2007). Denoting the scaled residual in the equilibrium conditions (15b) by

$$\mathbf{r}_h(q, \mathbf{v}, \dot{\mathbf{v}}, h\boldsymbol{\lambda}, t) := h\big(\mathbf{M}(q)\dot{\mathbf{v}} + \mathbf{g}(q, \mathbf{v}, t)\big) + \mathbf{B}^\top(q) \cdot h\boldsymbol{\lambda}\,,$$

we may rewrite the corrector equations (37) in the scaled and condensed form

$$\mathbf{0} = \boldsymbol{\Psi}_{n,h}(\boldsymbol{\xi}_{n+1}) := \begin{pmatrix} \mathbf{r}_h\big(q(\mathbf{\Delta q}_n), \mathbf{v}(\mathbf{\Delta q}_n), \dot{\mathbf{v}}(\mathbf{\Delta q}_n), h\boldsymbol{\lambda}_{n+1}, t_{n+1}\big) \\ \dfrac{1}{h}\, \boldsymbol{\Phi}\big(q(\mathbf{\Delta q}_n)\big) \end{pmatrix} \tag{57}$$

with $\boldsymbol{\xi}_{n+1} := \big((\mathbf{\Delta q}_n)^\top,\ h\boldsymbol{\lambda}_{n+1}^\top \big)^\top \in \mathbb{R}^{k+m}$ and

$$q_{n+1} = q(\mathbf{\Delta q}_n) := q_n \circ \exp(h\widetilde{\mathbf{\Delta q}}_n)\,, \tag{58a}$$

$$\mathbf{v}_{n+1} = \mathbf{v}(\mathbf{\Delta q}_n) := \frac{\gamma}{\beta}\mathbf{\Delta q}_n + (1 - \frac{\gamma}{\beta})\mathbf{v}_n + h(1 - \frac{\gamma}{2\beta})\mathbf{a}_n\,, \tag{58b}$$

$$\dot{\mathbf{v}}_{n+1} = \dot{\mathbf{v}}(\mathbf{\Delta q}_n) := \frac{1 - \alpha_m}{\beta(1 - \alpha_f)}\big(\frac{\mathbf{\Delta q}_n - \mathbf{v}_n}{h} - 0.5\mathbf{a}_n\big) + \frac{\mathbf{a}_n - \alpha_f \dot{\mathbf{v}}_n}{1 - \alpha_f}\,. \tag{58c}$$

The Newton–Raphson iteration

$$\boldsymbol{\xi}_{n+1}^{(k+1)} = \boldsymbol{\xi}_{n+1}^{(k)} + \mathbf{\Delta}\boldsymbol{\xi}_{n+1}^{(k)} \quad \text{with} \quad \frac{\partial \boldsymbol{\Psi}_{n,h}}{\partial \boldsymbol{\xi}}(\boldsymbol{\xi}_{n+1}^{(k)})\, \mathbf{\Delta}\boldsymbol{\xi}_{n+1}^{(k)} = -\boldsymbol{\Psi}_{n,h}(\boldsymbol{\xi}_{n+1}^{(k)}) \tag{59}$$

may be started, e.g., with the initial guess $\boldsymbol{\xi}_{n+1}^{(0)} = \big(\mathbf{v}_n^\top + 0.5h\mathbf{a}_n^\top,\ h\boldsymbol{\lambda}_n^\top\big)^\top$, see also (Brüls et al. 2012, Table 1) for an alternative definition of $\boldsymbol{\xi}_{n+1}^{(0)}$ and for a more detailed description of the full algorithm. The iteration matrix $\partial\boldsymbol{\Psi}_{n,h}/\partial\boldsymbol{\xi}$ has a 2 × 2-block structure

$$\frac{\partial\boldsymbol{\Psi}_{n,h}}{\partial\boldsymbol{\xi}} = \begin{pmatrix} \dfrac{1-\alpha_m}{\beta(1-\alpha_f)}\mathbf{M} + h\,\dfrac{\gamma}{\beta}\mathbf{D} + h^2\,\mathbf{K}\,\mathbf{T} & \mathbf{B}^\top \\ \mathbf{B}\,\mathbf{T} & \mathbf{0} \end{pmatrix} \tag{60}$$

with mass matrix $\mathbf{M} = \mathbf{M}\big(q(\boldsymbol{\Delta}\mathbf{q}_n)\big) \in \mathbb{R}^{k\times k}$, damping matrix

$$\mathbf{D} = \frac{\partial\mathbf{g}}{\partial\mathbf{v}}\big(q(\boldsymbol{\Delta}\mathbf{q}_n), \mathbf{v}(\boldsymbol{\Delta}\mathbf{q}_n), t_{n+1}\big) \in \mathbb{R}^{k\times k}\,,$$

constraint matrix $\mathbf{B} = \mathbf{B}\big(q(\boldsymbol{\Delta}\mathbf{q}_n)\big) \in \mathbb{R}^{m\times k}$ and the tangent operator $\mathbf{T} = \mathbf{T}(h\boldsymbol{\Delta}\mathbf{q}_n) \in \mathbb{R}^{k\times k}$ that results from the derivative of the exponential map in (58a), see Corollary 2.7. The stiffness matrix $\mathbf{K} = \mathbf{K}(q, \mathbf{v}, \dot{\mathbf{v}}, \boldsymbol{\lambda}, t) \in \mathbb{R}^{k\times k}$ represents the partial derivatives of the equilibrium equations (15b) w.r.t. $q \in G$ in the sense that

$$D_q\big(\mathbf{M}(q)\dot{\mathbf{v}} + \mathbf{g}(q, \mathbf{v}, t) + \mathbf{B}^\top(q)\boldsymbol{\lambda}\big) \cdot \big(DL_q(e) \cdot \widetilde{\mathbf{w}}\big) = \mathbf{K}(q, \mathbf{v}, \dot{\mathbf{v}}, \boldsymbol{\lambda}, t)\,\mathbf{w}$$

for all $\mathbf{w} \in \mathbb{R}^k$. It is evaluated at $q = q(\boldsymbol{\Delta}\mathbf{q}_n)$, $\mathbf{v} = \mathbf{v}(\boldsymbol{\Delta}\mathbf{q}_n)$, $\dot{\mathbf{v}} = \dot{\mathbf{v}}(\boldsymbol{\Delta}\mathbf{q}_n)$, $\boldsymbol{\lambda} = \boldsymbol{\lambda}_{n+1}$ and $t = t_{n+1}$.

The algorithmic parameters α_m, α_f and β in (37) satisfy $\alpha_m \neq 1, \alpha_f \neq 1$ and $\beta \neq 0$ since otherwise q_{n+1} would be independent of $\dot{\mathbf{v}}_{n+1}$ (and therefore also independent of the equilibrium equations (37e) at $t = t_{n+1}$). Hence, the iteration matrix $\partial\boldsymbol{\Psi}_{n,h}/\partial\boldsymbol{\xi}$ in (60) is non-singular for sufficiently small time step sizes h if the mass matrix $\mathbf{M}(q)$ is symmetric, positive definite and the constraint matrix $\mathbf{B}(q)$ has full rank (note, that $\mathbf{T}(h\boldsymbol{\Delta}\mathbf{q}_n) = \mathbf{I}_k + \mathcal{O}(h)$).

For sufficiently small time step sizes $h > 0$, the convergence of the Newton–Raphson iteration (59) may always be guaranteed under reasonable assumptions on $q_n, \mathbf{v}_n$:

Lemma 3.3 *If $\alpha_m \neq 1, \alpha_f \neq 1, \beta \neq 0$ and the numerical solution satisfies at $t = t_n$ the (hidden) constraints with residuals $\|\boldsymbol{\Phi}(q_n)\| \leq \gamma_0 h$ and $\|\mathbf{B}(q_n)\mathbf{v}_n\| \leq \gamma_0$ and a sufficiently small constant $\gamma_0 > 0$ then the generalized-α method (37) is well defined since the Newton–Raphson iteration (59) with initial guess $\boldsymbol{\xi}_{n+1}^{(0)} = (\mathbf{v}_n^\top, \mathbf{0}^\top)^\top + \mathcal{O}(h)$ converges for all sufficiently small time step sizes $h > 0$ to a locally uniquely defined solution of (57) with $\boldsymbol{\xi}_{n+1} = \boldsymbol{\xi}_{n+1}^{(0)} + \mathcal{O}(h) + \mathcal{O}(\gamma_0)$.*

Proof The assumptions on $\boldsymbol{\Phi}(q_n)$, $\mathbf{B}(q_n)\mathbf{v}_n$ and $\boldsymbol{\xi}_{n+1}^{(0)}$ are sufficient to prove $\boldsymbol{\Psi}_{n,h}(\boldsymbol{\xi}_{n+1}^{(0)}) = \mathcal{O}(h) + \mathcal{O}(\gamma_0)$ since $\mathbf{r}_h = \mathcal{O}(h)$ by definition and $q(\boldsymbol{\Delta}\mathbf{q}_n^{(0)}) = q(\mathbf{v}_n) + \mathcal{O}(h) = q_n \circ \exp(h\widetilde{\mathbf{v}}_n) + \mathcal{O}(h)$ resulting in

$$\begin{aligned}\frac{1}{h}\,\|\boldsymbol{\Phi}\big(q(\boldsymbol{\Delta}\mathbf{q}_n^{(0)})\big)\| &= \frac{1}{h}\,\|\boldsymbol{\Phi}(q_n) + h\,\frac{\mathrm{d}}{\mathrm{d}h}\boldsymbol{\Phi}\big(q_n \circ \exp(h\widetilde{\mathbf{v}}_n)\big) + \mathcal{O}(h^2)\| \\ &\le \frac{1}{h}\,\|\boldsymbol{\Phi}(q_n)\| + \|\mathbf{B}\big(q_n \circ \exp(h\widetilde{\mathbf{v}}_n)\big)\mathbf{v}_n\| + \mathcal{O}(h) \\ &= \mathcal{O}(h) + \mathcal{O}(\gamma_0)\,,\end{aligned}$$

see (32). Therefore, the convergence of the Newton–Raphson iteration to a locally uniquely defined solution $\boldsymbol{\xi}_{n+1} = \boldsymbol{\xi}_{n+1}^{(0)} + \mathcal{O}(h) + \mathcal{O}(\gamma_0)$ of (57) is guaranteed whenever the constant $\gamma_0 > 0$ and the time step size $h > 0$ are sufficiently small (Kelley 1995). □

The corrector equations (56) of the Lie group integrator for the stabilized index-2 formulation (55) may be condensed as well replacing the left equations in (58b, 58c) by

$$\mathbf{v}_{n+1} = \mathbf{v}\big(\boldsymbol{\Delta}\mathbf{q}_n + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n\big)\,,\;\; \dot{\mathbf{v}}_{n+1} = \dot{\mathbf{v}}\big(\boldsymbol{\Delta}\mathbf{q}_n + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n\big)\,.$$

The resulting scaled system of nonlinear equations is given by

$$\mathbf{0} = \boldsymbol{\Psi}_{n,h}(\boldsymbol{\xi}_{n+1}) := \begin{pmatrix} \boldsymbol{\varrho}_h(\boldsymbol{\Delta}\mathbf{q}_n, h\boldsymbol{\lambda}_{n+1}, \boldsymbol{\eta}_n) \\ \dfrac{1}{h}\,\boldsymbol{\Phi}\big(q(\boldsymbol{\Delta}\mathbf{q}_n)\big) \\ \mathbf{B}\big(q(\boldsymbol{\Delta}\mathbf{q}_n)\big)\mathbf{v}\big(\boldsymbol{\Delta}\mathbf{q}_n + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n\big) \end{pmatrix} \tag{61}$$

with $\boldsymbol{\xi}_{n+1} := \big((\boldsymbol{\Delta}\mathbf{q}_n)^\top,\; h\boldsymbol{\lambda}_{n+1}^\top,\; \boldsymbol{\eta}_n^\top\big)^\top \in \mathbb{R}^{k+2m}$ and

$$\begin{aligned}&\boldsymbol{\varrho}_h(\boldsymbol{\Delta}\mathbf{q}_n, h\boldsymbol{\lambda}_{n+1}, \boldsymbol{\eta}_n) := \\ &\quad \mathbf{r}_h\big(q(\boldsymbol{\Delta}\mathbf{q}_n), \mathbf{v}\big(\boldsymbol{\Delta}\mathbf{q}_n + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n\big), \dot{\mathbf{v}}\big(\boldsymbol{\Delta}\mathbf{q}_n + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n\big), h\boldsymbol{\lambda}_{n+1}, t_{n+1}\big)\,.\end{aligned}$$

The scaling of equations and unknowns guarantees again that the condition number of the iteration matrix $\partial\boldsymbol{\Psi}_{n,h}/\partial\boldsymbol{\xi}$ is bounded for $h \to 0$. This iteration matrix has the 3×3-block structure

$$\frac{\partial\boldsymbol{\Psi}_{n,h}}{\partial\boldsymbol{\xi}} = \begin{pmatrix} \mathbf{M}^* + h^2\,\mathbf{K}\,\mathbf{T} & \mathbf{B}^\top & \mathbf{M}^*\,\mathbf{B}^\top(q_n) \\ \mathbf{B}\,\mathbf{T} & \mathbf{0} & \mathbf{0} \\ \dfrac{\gamma}{\beta}\mathbf{B} + h\,\mathbf{Z} & \mathbf{0} & \dfrac{\gamma}{\beta}\,\mathbf{B}\mathbf{B}^\top(q_n) \end{pmatrix} \tag{62}$$

with

$$\mathbf{M}^* := \frac{1-\alpha_m}{\beta(1-\alpha_f)}\mathbf{M} + h\,\frac{\gamma}{\beta}\mathbf{D}$$

and a matrix $\mathbf{Z} \in \mathbb{R}^{k\times k}$ that represents $\big(\partial/\partial(\boldsymbol{\Delta}\mathbf{q}_n)\big)\mathbf{B}\big(q(\boldsymbol{\Delta}\mathbf{q}_n)\big)\mathbf{v}$ in the sense that

$$\mathbf{Z}\mathbf{w} = \mathbf{Z}\big(q(\boldsymbol{\Delta}\mathbf{q}_n)\big)\big(\mathbf{v}\big(\boldsymbol{\Delta}\mathbf{q}_n + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n\big), \mathbf{T}(h\boldsymbol{\Delta}\mathbf{q}_n)\mathbf{w}\big)\,,\;\; (\,\mathbf{w} \in \mathbb{R}^k\,)\,,$$

see (17). Using the formal decomposition

$$\frac{\partial \mathbf{\Psi}_{n,h}}{\partial \boldsymbol{\xi}} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} & \mathbf{M}^* \, \mathbf{B}^\top(q_n) \\ \mathbf{0} & \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \dfrac{\gamma}{\beta}\mathbf{I}_m & \dfrac{\gamma}{\beta}\mathbf{B}\mathbf{B}^\top(q_n) \end{pmatrix} \begin{pmatrix} \mathbf{M}^* + \mathcal{O}(h) & \mathbf{B}^\top & \mathbf{0} \\ \mathbf{B}\,\mathbf{T} & \mathbf{0} & \mathbf{0} \\ \mathcal{O}(h) & \mathbf{0} & \mathbf{I}_m \end{pmatrix},$$

see (62), we may verify that the iteration matrix is non-singular if $h > 0$ is sufficiently small. With the additional assumptions $\gamma \neq 0$ and $\boldsymbol{\eta}_n^{(0)} = \mathcal{O}(h)$, Lemma 3.3 applies also to the Lie group integrator (56) for the stabilized index-2 formulation. The method is well defined and the corresponding condensed system (61) may be solved by the Newton–Raphson method (59).

In the practical implementation of implicit ODE/DAE time integration methods, the Jacobian $(\partial \mathbf{\Psi}_{n,h}/\partial \boldsymbol{\xi})(\boldsymbol{\xi}_{n+1}^{(k)})$ in the Newton–Raphson step (59) is substituted by an approximation that is kept constant during integration as long as possible, see, e.g., (Brenan et al. 1996, Sect. 5.2.2). In (Brüls et al. 2011), the influence of different Lie group formulations on the number of Jacobian updates was studied by numerical tests for the Lie group integrator (37). A very small number of Jacobian evaluations were observed for equations of motion like (22) that are characterized by a constant mass matrix **M** and a constant constraint Jacobian **B**, see also Lemma 3.5 below.

If the generalized-α integrators (37) and (56) are applied to non-stiff systems and the time step size h is sufficiently small, then we may neglect in (60) and (62) the terms $h\gamma\mathbf{D}/\beta$, $h^2\mathbf{KT}$ and $h\mathbf{Z}$. For the numerical tests in Sects. 3.3 and 3.4, this simplified Newton–Raphson method was combined with a damping strategy based on Armijo line search, see (Kelley 1995). Convergence problems in the corrector iteration were observed for just one simulation scenario (integrator (37) for the heavy top benchmark, $G = \mathrm{SO}(3) \times \mathbb{R}^3$, $h = 4.0 \times 10^{-3}$, see the left plots of Figs. 6 and 8). Here, we had to take into account a difference approximation of the term $h\gamma\mathbf{D}/\beta + h^2\mathbf{KT}$ in (60).

3.6 Constraint Residuals

Both generalized-α integrators (37) and (56) satisfy by construction the holonomic constraints (15c) at the level of position coordinates: $\mathbf{\Phi}(q_n) = \mathbf{0}$, $(\, n > 0\,)$. For the stabilized index-2 integrator (56), the hidden constraints (16) at velocity level are satisfied as well: $\mathbf{B}(q_n)\mathbf{v}_n = \mathbf{0}$, $(\, n > 0\,)$, see (56g). For the index-3 integrator (37), these residuals $\mathbf{B}(q_n)\mathbf{v}_n$ remain in general in the size of global discretization errors since $\mathbf{B}(q(t))\mathbf{v}(t) \equiv \mathbf{0}$. For some problem classes, the constraint residuals $\mathbf{B}(q_n)\mathbf{v}_n$ vanish, however, also for the index-3 integrator (37). Therefore, both integrators (37) and (56) define in that case one and the same numerical solution $(q_n, \mathbf{v}_n, \dot{\mathbf{v}}_n, \mathbf{a}_n, \boldsymbol{\lambda}_n)$ with auxiliary variables $\boldsymbol{\eta}_n = \mathbf{0}$, $(\, n \geq 0\,)$. In a practical implementation, the numerical solutions will coincide up to round-off errors and errors that are caused by stopping the Newton–Raphson iteration after a finite number of iteration steps.

In the present section, we show that the numerical solution of the index-3 integrator (37) will always satisfy the hidden constraints (16) at the level of velocity coordinates if the constraint Jacobian $\mathbf{B}$ is constant (Lemma 3.4). In Lemma 3.5, this result is extended to a special problem class in SE(3) with $\mathbf{B}(q) = \text{const}$ on the constraint manifold $\mathfrak{M} = \{\, q \in G \,:\, \boldsymbol{\Phi}(q) = \mathbf{0}\,\}$. This analysis gives the formal proof for the numerical test results in the right plot of Fig. 10 that were obtained for the heavy top benchmark in configuration space $G = \text{SE}(3)$.

Improved error estimates for certain configuration spaces are a topic of active current research on Lie group time integration methods, see also the recently published results of Müller and Terze (2014a, b).

Lemma 3.4 *Consider equations of motion (15) with constant constraint Jacobian* $\mathbf{B}$ *in the hidden constraints (16) at velocity level.*

(a) *For this problem class, the curvature term* $\mathbf{Z}(q)\big(\mathbf{v}, \mathbf{v}\big)$ *in the hidden constraints (18) at acceleration level vanishes identically.*
(b) *If* $\mathbf{B} = \text{const}$ *and the starting values* q_0, $\mathbf{v}_0$, $\mathbf{a}_0$ *are consistent (* $\mathbf{0} = \boldsymbol{\Phi}(q_0) = \mathbf{B}\mathbf{v}_0 = \mathbf{B}\mathbf{a}_0$ *) then the numerical solution* $(q_n, \mathbf{v}_n, \dot{\mathbf{v}}_n, \mathbf{a}_n, \boldsymbol{\lambda}_n)$ *of the generalized-*α *method (37) satisfies for all* $n \geq 0$ *both the holonomic constraints (15c) at position level and the hidden constraints (16) at velocity level:* $\boldsymbol{\Phi}(q_n) = \mathbf{B}\mathbf{v}_n = \mathbf{0}$.

Proof (a) The time derivative of hidden constraints (16) with $\mathbf{B} = \text{const}$ is given by $\mathbf{0} = \mathbf{B}\dot{\mathbf{v}}(t)$. Comparing this expression with the hidden constraints (18), we get $\mathbf{Z}(q)\big(\mathbf{v}, \mathbf{v}\big) = \mathbf{0}$.

(b) Because of $\boldsymbol{\Phi}(q_0) = \mathbf{0}$ and (37f), the numerical solution q_n satisfies the holonomic constraints (15c) for all $n \geq 0$. To prove $\mathbf{B}\mathbf{v}_n = \mathbf{B}\mathbf{a}_n = \mathbf{0}$ by induction, we observe that $\boldsymbol{\Phi}(q_{n+1}) = \boldsymbol{\Phi}(q_n) = \mathbf{0}$ and $q_{n+1} = q_n \circ \exp(h\widetilde{\boldsymbol{\Delta}\mathbf{q}}_n)$, see (37a), imply $\boldsymbol{\Psi}(1) = \boldsymbol{\Psi}(0) = \mathbf{0}$ for the continuously differentiable function $\boldsymbol{\Psi} : [0, 1] \to \mathbb{R}^m$, $\vartheta \mapsto \boldsymbol{\Phi}\big(q_n \circ \exp(\vartheta h\widetilde{\boldsymbol{\Delta}\mathbf{q}}_n)\big)$. Therefore,

$$\begin{aligned} \mathbf{0} = \frac{\boldsymbol{\Psi}(1) - \boldsymbol{\Psi}(0)}{h} &= \frac{1}{h}\int_0^1 \frac{\mathrm{d}\boldsymbol{\Phi}}{\mathrm{d}\vartheta}\big(q_n \circ \exp(\vartheta h\widetilde{\boldsymbol{\Delta}\mathbf{q}}_n)\big)\,\mathrm{d}\vartheta \\ &= \int_0^1 \mathbf{B}\big(q_n \circ \exp(\vartheta h\widetilde{\boldsymbol{\Delta}\mathbf{q}}_n)\big)\boldsymbol{\Delta}\mathbf{q}_n\,\mathrm{d}\vartheta\,, \end{aligned} \tag{63}$$

see (14) and (32). If $\mathbf{B} = \text{const}$, then the integrand in (63) is constant as well resulting in $\mathbf{B}\boldsymbol{\Delta}\mathbf{q}_n = \mathbf{0}$. We get $\mathbf{B}\mathbf{a}_{n+1} = \mathbf{0}$ (if $\mathbf{B}\mathbf{v}_n = \mathbf{B}\mathbf{a}_n = \mathbf{0}$) from left multiplication of (37b) by matrix $\mathbf{B}$ and obtain finally $\mathbf{B}\mathbf{v}_{n+1} = \mathbf{0}$ multiplying also the velocity update (37c) from the left by the (constant) constraint Jacobian $\mathbf{B}$. □

Lemma 3.5 *Consider a rigid body with configuration space* SE(3) *and holonomic constraints (15c) of the form*

$$\mathbf{0} = \boldsymbol{\Phi}(q) = \boldsymbol{\Phi}\big((\mathbf{R}, \mathbf{x})_{\mathrm{SE}(3)}\big) = \mathbf{X} - \mathbf{R}^\top \mathbf{x} \tag{64}$$

with a constant vector $\mathbf{X} \in \mathbb{R}^3$.

(a) Along any solution $q(t)$ *of the constrained equations of motion (15) matrix* $\mathbf{B}\big(q(t)\big)$ *is constant and the curvature term* $\mathbf{Z}\big(q(t)\big)\big(\mathbf{v}(t), \mathbf{v}(t)\big)$ *vanishes identically.*

(b) If the generalized-α *method (37) is applied with consistent starting values (* $\mathbf{0} = \boldsymbol{\Phi}(q_0) = \mathbf{B}(q_0)\mathbf{v}_0 = \mathbf{B}(q_0)\mathbf{a}_0$*) and with sufficiently small time step size* $h > 0$ *to equations of motion (15) in* SE(3) *with holonomic constraints (64) then the numerical solution satisfies both the holonomic constraints at position level and the hidden constraints at velocity level:* $\boldsymbol{\Phi}(q_n) = \mathbf{B}(q_n)\mathbf{v}_n = \mathbf{0}$*, (* $n \geq 0$ *).*

Proof (a) Straightforward differentiation of constraint (64) shows

$$\begin{aligned}\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Phi}(q(t)) &= -\dot{\mathbf{R}}^\top \mathbf{x} - \mathbf{R}^\top \dot{\mathbf{x}} = -(\mathbf{R}\widetilde{\boldsymbol{\Omega}})^\top \mathbf{x} - \mathbf{R}^\top \mathbf{R}\mathbf{U} \\ &= -\widetilde{\boldsymbol{\Omega}}^\top \mathbf{R}^\top \mathbf{x} - \mathbf{U} = \widetilde{\boldsymbol{\Omega}}\mathbf{R}^\top \mathbf{x} - \mathbf{U} = -\widetilde{\mathbf{R}^\top \mathbf{x}}\,\boldsymbol{\Omega} - \mathbf{U} = \mathbf{B}(q)\mathbf{v}\end{aligned}$$

with $q = (\mathbf{R}, \mathbf{x})_{\mathrm{SE}(3)} \in \mathrm{SE}(3)$ and $\mathbf{v} = (\,\boldsymbol{\Omega}^\top,\ \mathbf{U}^\top)^\top \in \mathbb{R}^6$. On the constraint manifold, we have $\mathbf{R}^\top \mathbf{x} = \mathbf{X}$, see (64), and the constraint Jacobian $\mathbf{B}(q)$ is constant: $\mathbf{B}\big((\mathbf{R}, \mathbf{x})_{\mathrm{SE}(3)}\big) = \mathbf{B}^{\mathbf{X}} := \big(-\widetilde{\mathbf{X}}\ \ -\mathbf{I}_3\,\big)$. Therefore, the hidden constraints (16) and (18) are given by $\mathbf{B}^{\mathbf{X}}\mathbf{v}(t) = \mathbf{0}$ and $\mathbf{B}^{\mathbf{X}}\dot{\mathbf{v}}(t) = \mathbf{0}$ with $\mathbf{Z}\big(q(t)\big)\big(\mathbf{v}(t), \mathbf{v}(t)\big) \equiv \mathbf{0}$ along any solution $\big(q(t), \mathbf{v}(t)\big)$.

(b) This part of the proof is substantially more technical than the corresponding proof of Lemma 3.4(b) since $\mathbf{B}(q)$ is not constant beyond the constraint manifold $\mathfrak{M}$ and there is no straightforward way to prove that in (63) the argument $q_n \circ \exp(\vartheta h \widetilde{\boldsymbol{\Delta}\mathbf{q}}_n)$ of $\mathbf{B}$ will remain in $\mathfrak{M}$ for $\vartheta \in (0, 1)$.

In SE(3), the position update formula $q_{n+1} = q_n \circ \exp(h\widetilde{\boldsymbol{\Delta}\mathbf{q}}_n)$ gets the form

$$\mathbf{R}_{n+1} = \mathbf{R}_n \exp_{\mathrm{SO}(3)}(h\widetilde{\boldsymbol{\Delta}\mathbf{R}}_n)\,, \quad \mathbf{x}_{n+1} = \mathbf{x}_n + h\mathbf{R}_n \mathbf{T}^\top_{\mathrm{SO}(3)}(h\boldsymbol{\Delta}\mathbf{R}_n)\boldsymbol{\Delta}\mathbf{x}_n$$

with $\boldsymbol{\Delta}\mathbf{q}_n = (\,\boldsymbol{\Delta}\mathbf{R}_n^\top,\ \boldsymbol{\Delta}\mathbf{x}_n^\top)^\top$, see Example 2.1(a). Because of $\boldsymbol{\Phi}(q_0) = \mathbf{0}$ and $\boldsymbol{\Phi}(q_{n+1}) = \mathbf{0}$, ($n \geq 0$), see (37f), we get $\mathbf{R}_n^\top \mathbf{x}_n - \mathbf{R}_{n+1}^\top \mathbf{x}_{n+1} = \mathbf{X} - \mathbf{X} = \mathbf{0}$, see (64), and

$$\begin{aligned}\mathbf{0} &= \exp_{\mathrm{SO}(3)}(h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n})\,\frac{\mathbf{R}_n^\top\mathbf{x}_n-\mathbf{R}_{n+1}^\top\mathbf{x}_{n+1}}{h}\\ &= \frac{\exp_{\mathrm{SO}(3)}(h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n})\mathbf{R}_n^\top\mathbf{x}_n-\mathbf{R}_n^\top\Big(\mathbf{x}_n+h\mathbf{R}_n\mathbf{T}_{\mathrm{SO}(3)}^\top(h\boldsymbol{\Delta}\mathbf{R}_n)\boldsymbol{\Delta}\mathbf{x}_n\Big)}{h}\\ &= \frac{\exp_{\mathrm{SO}(3)}(h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n})-\mathbf{I}_3}{h}\,\mathbf{R}_n^\top\mathbf{x}_n-\mathbf{T}_{\mathrm{SO}(3)}^\top(h\boldsymbol{\Delta}\mathbf{R}_n)\boldsymbol{\Delta}\mathbf{x}_n \qquad (65)\end{aligned}$$

with

$$\begin{aligned}\exp_{\mathrm{SO}(3)}(h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n})-\mathbf{I}_3 &= \sum_{i=1}^{\infty}\frac{1}{i!}\big(h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}\big)^i = h\sum_{i=0}^{\infty}\frac{1}{(i+1)!}\big(h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}\big)^i\,\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}\\ &= h\sum_{i=0}^{\infty}\frac{(-1)^i}{(i+1)!}\big(-h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}\big)^i\,\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}\,.\end{aligned}$$

In SO(3), the $\widetilde{(\bullet)}$ operator maps $\boldsymbol{\Delta}\mathbf{R}_n \in \mathbb{R}^3$ to the skew symmetric matrix $\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}$, see (2), and we have $\widehat{\boldsymbol{\Delta}\mathbf{R}_n} = \widetilde{\boldsymbol{\Delta}\mathbf{R}_n}$, see Remark 2.8(b). Therefore, $-\widetilde{\boldsymbol{\Delta}\mathbf{R}_n} = (\widetilde{\boldsymbol{\Delta}\mathbf{R}_n})^\top = (\widehat{\boldsymbol{\Delta}\mathbf{R}_n})^\top$ and the series expansion (30) proves

$$\exp_{\mathrm{SO}(3)}(h\widetilde{\boldsymbol{\Delta}\mathbf{R}_n})-\mathbf{I}_3 = h\left(\mathbf{T}_{\mathrm{SO}(3)}(h\boldsymbol{\Delta}\mathbf{R}_n)\right)^\top\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}\,.$$

Inserting this expression in (65), we get

$$\mathbf{0} = \mathbf{T}_{\mathrm{SO}(3)}^\top(h\boldsymbol{\Delta}\mathbf{R}_n)\big(\widetilde{\boldsymbol{\Delta}\mathbf{R}_n}(\mathbf{R}_n^\top\mathbf{x}_n)-\boldsymbol{\Delta}\mathbf{x}_n\big)$$

and therefore also

$$\mathbf{0} = \widetilde{\boldsymbol{\Delta}\mathbf{R}_n}(\mathbf{R}_n^\top\mathbf{x}_n)-\boldsymbol{\Delta}\mathbf{x}_n = -\widetilde{\mathbf{R}_n^\top\mathbf{x}_n}\,\boldsymbol{\Delta}\mathbf{R}_n-\boldsymbol{\Delta}\mathbf{x}_n = \mathbf{B}(q_n)\boldsymbol{\Delta}\mathbf{q}_n$$

since the tangent operator $\mathbf{T}_{\mathrm{SO}(3)}(h\boldsymbol{\Delta}\mathbf{R}_n) = \mathbf{I}_3 + \mathcal{O}(h)$ is non-singular for sufficiently small time step sizes $h > 0$. Now, the proof may be completed following line by line the proof of Lemma 3.4(b) since $q_n \in \mathfrak{M}$ by construction and $\mathbf{B}(q)$ is constant on the constraint manifold, i.e., $\mathbf{B}(q_n) = \mathbf{B}^{\mathbf{X}} = \text{const}$. □

4 Convergence Analysis

The convergence of generalized-α time integration methods for nonlinear unconstrained systems in linear configuration spaces was studied by Erlicher et al. (2002) using an equivalent multi-step representation. In the DAE Lie group case, this analysis has to be extended to constrained systems in nonlinear configuration spaces with Lie group structure, see (Brüls et al. 2012). In the present section, we follow the direct convergence analysis for the generalized-α method in one-step form (37) that was developed in (Arnold et al. 2015) to study the convergence in long-term integration as well as in the transient phase in full detail.

4.1 Local Truncation Errors, Global Errors and Error Recursion

For unconstrained systems in linear spaces, the local truncation errors were introduced in (39), see Sect. 3.2 above. Since there are no discretization errors in the holonomic constraints (15c), see (37f), these definitions may be used as well in the constrained case.

For configuration spaces with Lie group structure, the definition of the local truncation error $\mathbf{l}_n^{\mathbf{q}}$ in (39a) has to be adapted to the Lie group setting. In the Lie algebra approach to error analysis of Lie group time integration methods, we follow the proposal of Wensch (2001) to define local and global errors by elements of the corresponding Lie algebra, see also (Orel 2010):

Definition 4.1 For the solution components $q \in G$, the *local truncation error* $\widetilde{\mathbf{l}}_n^q \in \mathfrak{g}$ of the generalized-α Lie group method (37) is defined by

$$q(t_{n+1}) = q(t_n) \circ \exp(h\widetilde{\mathbf{\Delta q}}(t_n)) \circ \exp(\widetilde{\mathbf{l}}_n^q) \tag{66}$$

with $\mathbf{\Delta q}(t_n) := \mathbf{v}(t_n) + (0.5 - \beta)h\dot{\mathbf{v}}(t_n + \Delta_\alpha h) + \beta h\dot{\mathbf{v}}(t_{n+1} + \Delta_\alpha h)$.

To get an error estimate for $\widetilde{\mathbf{l}}_n^q$, we compare the asymptotic behaviour of $q(t_{n+1}) = q(t_n + h)$ and $q(t_n) \circ \exp(h\widetilde{\mathbf{\Delta q}}(t_n))$ for $h \to 0$. For any smooth function $\mathbf{v}(t)$, the flow of $\dot{q}(t) = DL_q(e) \cdot \widetilde{\mathbf{v}}(t)$ is locally represented by a smooth function $\widetilde{\boldsymbol{\nu}} : [-h_0, h_0] \times \mathbb{R} \times G \to \mathfrak{g}$:

$$q(t + h) = q(t) \circ \exp\bigl(h\widetilde{\boldsymbol{\nu}}(h; t, q(t))\bigr). \tag{67}$$

The asymptotic behaviour of $h\widetilde{\boldsymbol{\nu}}$ is characterized by the Magnus expansion

$$h\widetilde{\boldsymbol{\nu}}(h; t, q(t)) = h\widetilde{\mathbf{v}}(t) + \frac{h^2}{2}\dot{\widetilde{\mathbf{v}}}(t) + \frac{h^3}{6}\ddot{\widetilde{\mathbf{v}}}(t) + \frac{h^3}{12}[\widetilde{\mathbf{v}}(t), \dot{\widetilde{\mathbf{v}}}(t)] + \mathcal{O}(h^4)\,, \tag{68}$$

see (Hairer et al. 2006) and (Müller 2010). The matrix commutator $[\widetilde{\mathbf{v}}, \dot{\widetilde{\mathbf{v}}}]$ vanishes identically in linear spaces, see Sect. 2.5. In the Lie group setting, it introduces an additional local error term if the arguments $\widetilde{\mathbf{v}}(t)$ and $\dot{\widetilde{\mathbf{v}}}(t)$ do not commute, see Lemma 4.2 below.

Inserting (67) with $t = t_n$ into the (implicit) definition of $\widetilde{\mathbf{l}}_n^q$, see (66), we get $q(t_n) \circ \exp\bigl(h\widetilde{\boldsymbol{\nu}}(h; t_n, q(t_n))\bigr) = q(t_n) \circ \exp(h\widetilde{\mathbf{\Delta q}}(t_n)) \circ \exp(\widetilde{\mathbf{l}}_n^q)$. Therefore, the term $\exp(\widetilde{\mathbf{l}}_n^q)$ may be expressed as product of matrix exponentials:

$$\exp(\widetilde{\mathbf{l}}_n^q) = \exp(-h\widetilde{\mathbf{\Delta q}}(t_n)) \circ \exp\big(h\widetilde{\boldsymbol{\nu}}(h; t_n, q(t_n))\big)\,.$$

In (Arnold et al. 2015, Lemma 1), we used the Baker–Campbell–Hausdorff formula to show that $\widetilde{\mathbf{l}}_n^q$ and $h\big(\widetilde{\boldsymbol{\nu}}(h; t_n, q(t_n)) - \widetilde{\mathbf{\Delta q}}(t_n)\big)$ coincide up to higher order terms, see also Lemma 2.5. Comparing the Magnus expansion (68) with the Taylor expansion of $\widetilde{\mathbf{\Delta q}}(t_n)$, we get

Lemma 4.2 *With $\Delta_\alpha := \alpha_m - \alpha_f$ and $C_q := (1 - 6\beta - 3\Delta_\alpha)/6$, the local truncation error $\widetilde{\mathbf{l}}_n^q$ is given by*

$$\widetilde{\mathbf{l}}_n^q = C_q h^3 \widetilde{\ddot{\mathbf{v}}}(t_n) + h^3[\widetilde{\mathbf{v}}(t_n), \widetilde{\dot{\mathbf{v}}}(t_n)]/12 + \mathcal{O}(h^4)\,. \tag{69}$$

If the parameters γ, α_m, α_f satisfy the order condition (41) then the local truncation errors are bounded by

$$\|\mathbf{l}_n^q\| = \mathcal{O}(h^3)\,, \quad \|\mathbf{l}_{n+1}^q - \mathbf{l}_n^q\| = \mathcal{O}(h^4)\,, \quad \|\mathbf{l}_n^{\mathbf{v}}\| = \mathcal{O}(h^3)\,, \quad \|\mathbf{l}_n^{\mathbf{a}}\| = \mathcal{O}(h^2)\,. \tag{70}$$

The linear relations between $\mathbf{v}_n$, $\mathbf{a}_n$ and $\dot{\mathbf{v}}_n$ in (37) result in linear relations for the corresponding global errors. Here and in the following we will always assume that the algorithmic parameters γ, α_m and α_f satisfy the order condition (41) and the local truncation errors are bounded by (70).

Lemma 4.3 *Consider global errors $\mathbf{e}_n^{\mathbf{a}}$ with $\dot{\mathbf{v}}(t_n + \Delta_\alpha h) = \mathbf{a}_n + \mathbf{e}_n^{\mathbf{a}}$ and use $(\bullet)(t_n) = (\bullet)_n + \mathbf{e}_n^{(\bullet)}$ to define $\mathbf{e}_n^{(\bullet)}$ for all remaining solution components being elements of linear spaces. The order condition (41) implies*

$$\mathbf{e}_{n+1}^{\mathbf{v}} = \mathbf{e}_n^{\mathbf{v}} + (1-\gamma)h\mathbf{e}_n^{\mathbf{a}} + \gamma h\mathbf{e}_{n+1}^{\mathbf{a}} + \mathcal{O}(h^3)\,, \tag{71a}$$

$$(1-\alpha_m)\mathbf{e}_{n+1}^{\mathbf{a}} + \alpha_m\mathbf{e}_n^{\mathbf{a}} = (1-\alpha_f)\mathbf{e}_{n+1}^{\dot{\mathbf{v}}} + \alpha_f\mathbf{e}_n^{\dot{\mathbf{v}}} + \mathcal{O}(h^2)\,. \tag{71b}$$

For linear configuration spaces G, the global error in $\mathbf{q}$ is given by $\mathbf{q}(t_n) = \mathbf{q}_n + \mathbf{e}_n^{\mathbf{q}}$. In the nonlinear case, we take into account the Lie group structure of the configuration space G and consider global errors $\widetilde{\mathbf{e}}_n^q$ being elements of the corresponding Lie algebra $\mathfrak{g}$:

$$q(t_n) = q_n \circ \exp(\widetilde{\mathbf{e}}_n^q)\,. \tag{72}$$

This definition is compatible with the classical definition of $\mathbf{e}_n^{\mathbf{q}} \in \mathbb{R}^k$ if the configuration space G is linear.

The position update (37a) and the definition (66) of the local error $\widetilde{\mathbf{l}}_n^q$ yield a global error recursion for $\widetilde{\mathbf{e}}_n^q$ in terms of matrix exponentials:

$$\begin{aligned}\exp(\widetilde{\mathbf{e}}_{n+1}^q) &= (q_{n+1})^{-1} \circ q(t_{n+1})\\ &= \exp(-h\widetilde{\mathbf{\Delta q}}_n) \circ \underbrace{(q_n)^{-1} \circ q(t_n)}_{=\exp(\widetilde{\mathbf{e}}_n^q)} \circ \exp(h\widetilde{\mathbf{\Delta q}}(t_n)) \circ \exp(\widetilde{\mathbf{l}}_n^q)\,.\end{aligned}$$

This product of matrix exponentials may be studied by repeated application of the Baker–Campbell–Hausdorff formula using Lemma 2.5. Omitting all technical details, we get

Lemma 4.4 (Arnold et al. 2015, Lemma 2) *The global errors* $\mathbf{e}_n^q$ *satisfy*

$$\mathbf{e}_{n+1}^q = \mathbf{e}_n^q + h\,\boldsymbol{\Delta}_h \mathbf{e}_n^q \tag{73}$$

with

$$\begin{aligned}\boldsymbol{\Delta}_h \widetilde{\mathbf{e}}_n^q = \widetilde{\mathbf{e}}_n^{\mathbf{v}} + (0.5-\beta)h\widetilde{\mathbf{e}}_n^{\mathbf{a}} + \beta h\widetilde{\mathbf{e}}_{n+1}^{\mathbf{a}} + [\widetilde{\mathbf{e}}_n^q, \widetilde{\mathbf{v}}(t_n)] + \frac{1}{h}\widetilde{\mathbf{l}}_n^q + \\ + \mathcal{O}(h)(\varepsilon_n + h\|\mathbf{e}_{n+1}^{\mathbf{a}}\|)\end{aligned} \tag{74}$$

and the notation

$$\varepsilon_n := \|\mathbf{e}_n^q\| + \|\mathbf{e}_n^{\mathbf{v}}\| + h\|\mathbf{e}_n^{\mathbf{a}}\| + h\|\mathbf{e}_n^{\boldsymbol{\lambda}}\| \tag{75}$$

that is used to summarize higher order error terms in compact form. In particular, Eqs. (73) and (74) and the local error estimate (69) imply

$$\mathbf{e}_{n+1}^q = \mathbf{e}_n^q + \mathcal{O}(h)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^3)\,, \tag{76a}$$

$$\|\boldsymbol{\Delta}_h \mathbf{e}_n^q\| \le \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2)\,. \tag{76b}$$

Error estimates like the ones in Lemma 4.4 are valid if the numerical solution remains in a small neighbourhood of the analytical one. More precisely, we suppose that there are positive constants h_0 and C and a sufficiently small constant $\gamma_0 > 0$ such that

$$\|\mathbf{e}_r^q\| \le Ch\,, \quad \|\mathbf{e}_r^{\mathbf{v}}\| + \|\mathbf{e}_r^{\mathbf{a}}\| + \|\mathbf{e}_r^{\boldsymbol{\lambda}}\| \le \gamma_0 \tag{77}$$

is satisfied for all $h \in (0, h_0]$ and all r with $t_0 + rh \in [t_0, t_{\text{end}}]$. This technical assumption may be verified using the results of the convergence analysis in Sect. 4.3 below, see (Hairer and Wanner 1996, Theorem VII.3.5) and the slightly more detailed discussion in (Arnold et al. 2015, Sect. 3.1).

Linearizing the equilibrium conditions (37e), we may estimate $\mathbf{e}_n^{\dot{\mathbf{v}}}$ in terms of ε_n and $\mathbf{e}_n^{\boldsymbol{\lambda}}$:

Lemma 4.5 (Arnold et al. 2015, Lemma 3) *If the order condition (41) is satisfied then*

$$\mathbf{e}_n^{\dot{\mathbf{v}}} + \mathbf{e}_n^{\mathbf{M}^{-1}\mathbf{B}^\top\boldsymbol{\lambda}} = \mathcal{O}(1)\varepsilon_n\,, \quad \|\mathbf{e}_n^{\dot{\mathbf{v}}}\| = \mathcal{O}(1)(\varepsilon_n + \|\mathbf{e}_n^{\boldsymbol{\lambda}}\|)\,, \tag{78a}$$

$$\mathbf{e}_{n+1}^{\dot{\mathbf{v}}} + \mathbf{e}_{n+1}^{\mathbf{M}^{-1}\mathbf{B}^\top\boldsymbol{\lambda}} = \mathcal{O}(1)\varepsilon_n + \mathcal{O}(h)(\|\mathbf{e}_{n+1}^{\mathbf{a}}\| + \|\mathbf{e}_{n+1}^{\boldsymbol{\lambda}}\|) + \mathcal{O}(h^3)\,. \tag{78b}$$

Here we used the notation $\mathbf{e}_n^{(\mathbf{C}\,\bullet)} := \mathbf{C}(q(t_n), \mathbf{v}(t_n), \boldsymbol{\lambda}(t_n), t_n)\mathbf{e}_n^{(\bullet)}$ *for matrix-valued functions* $\mathbf{C} = \mathbf{C}(q, \mathbf{v}, \boldsymbol{\lambda}, t)$.

Inserting (78) into the error estimate (71b), we get a coupled error recursion

$$\begin{aligned}(1-\alpha_m)\mathbf{e}^{\mathbf{a}}_{n+1} + \alpha_m \mathbf{e}^{\mathbf{a}}_n + (1-\alpha_f)\mathbf{e}^{\mathbf{M}^{-1}\mathbf{B}^\top\boldsymbol{\lambda}}_{n+1} + \alpha_f \mathbf{e}^{\mathbf{M}^{-1}\mathbf{B}^\top\boldsymbol{\lambda}}_n = \\ = \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2) \qquad (79)\end{aligned}$$

that has to be studied separately in tangential and normal direction of the constraint manifold $\mathfrak{M} := \{\, q \in G \;:\; \boldsymbol{\Phi}(q) = \mathbf{0}\,\}$ to get optimal error bounds, see (Hairer and Wanner 1996). The error component in tangential direction is obtained by multiplication with a matrix $\mathbf{P}(q)$ that projects into the tangential space $T_q\mathfrak{M} = \ker \mathbf{B}(q)$. Such a projector $\mathbf{P}(q)$ is given by

$$\mathbf{P}(q) := \mathbf{I} - [\mathbf{M}^{-1}\mathbf{B}^\top\mathbf{S}^{-1}\mathbf{B}](q) \;\text{ with }\; \mathbf{S}(q) := [\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^\top](q) \qquad (80)$$

since $\mathbf{PP} = \mathbf{P}$ and $\mathbf{BP} = \mathbf{B} - \mathbf{BM}^{-1}\mathbf{B}^\top\mathbf{S}^{-1}\mathbf{B} = \mathbf{B} - \mathbf{SS}^{-1}\mathbf{B} = \mathbf{0}$. Taking into account that this projector satisfies $\mathbf{PM}^{-1}\mathbf{B}^\top \equiv \mathbf{0}$, we get an optimal error recursion in tangential direction by left multiplication of (79) with matrix $\mathbf{P}\big(q(t_{n+1})\big)$. The error propagation in normal direction to the constrained manifold may be characterized multiplying (79) by $\mathbf{B}\big(q(t_{n+1})\big)$:

Lemma 4.6 (Arnold et al. 2015, Lemma 5) *The errors* $\mathbf{e}^{\mathbf{a}}_n$, $\mathbf{e}^{\boldsymbol{\lambda}}_n$ *satisfy*

$$(1-\alpha_m)\mathbf{e}^{\mathbf{Pa}}_{n+1} + \alpha_m \mathbf{e}^{\mathbf{Pa}}_n = \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2)\,, \qquad (81)$$

$$\begin{aligned}(1-\alpha_m)\mathbf{e}^{\mathbf{Ba}}_{n+1} + \alpha_m \mathbf{e}^{\mathbf{Ba}}_n + (1-\alpha_f)\mathbf{e}^{\mathbf{S}\boldsymbol{\lambda}}_{n+1} + \alpha_f \mathbf{e}^{\mathbf{S}\boldsymbol{\lambda}}_n = \\ = \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2) \qquad (82)\end{aligned}$$

and $\|\mathbf{e}^{\mathbf{a}}_n\| \le \|\mathbf{e}^{\mathbf{Pa}}_n\| + \|\mathbf{M}^{-1}\mathbf{B}^\top\mathbf{S}^{-1}\|\|\mathbf{e}^{\mathbf{Ba}}_n\| \le \mathcal{O}(1)(\|\mathbf{e}^{\mathbf{Pa}}_n\| + \|\mathbf{e}^{\mathbf{Ba}}_n\|)$.

Estimate (81) defines a one-step recursion for the tangential error component $\mathbf{e}^{\mathbf{Pa}}_n$ in terms of ε_n, ε_{n+1} and local errors $\mathcal{O}(h^2)$.

The most crucial part of the convergence analysis are recursive estimates for the error component $\mathbf{e}^{\mathbf{Ba}}_n$ in normal direction to the constrained manifold. Similar to the discussion in Sect. 3.2, we may scale the error recursion (71a) by the factor $1/h$ to get

$$(1-\gamma)\mathbf{e}^{\mathbf{Ba}}_n + \gamma \mathbf{e}^{\mathbf{Ba}}_{n+1} = \frac{\mathbf{e}^{\mathbf{Bv}}_{n+1} - \mathbf{e}^{\mathbf{Bv}}_n}{h} + \mathcal{O}(1)\varepsilon_n + \mathcal{O}(h^2)\,. \qquad (83)$$

The scaled error term $\mathbf{e}^{\mathbf{Bv}}_n/h$ in the right-hand side of (83) is studied considering error estimate (74) and its equivalent in $\mathbb{R}^k$. We get

$$\frac{1}{h}\Big(\mathbf{e}^{\mathbf{Bv}}_n + \frac{1}{h}\mathbf{B}\big(q(t_n)\big)\mathbf{l}^q_n\Big) = \mathbf{r}^{\mathbf{B}}_n - \mathbf{r}_h(t_n, \mathbf{e}^q_n) + \mathcal{O}(1)\varepsilon_n + \mathcal{O}(h)\|\mathbf{e}^{\mathbf{a}}_{n+1}\| \qquad (84)$$

with the vector

$$\mathbf{r}_n^{\mathbf{B}} := \frac{1}{h}\Big(\mathbf{B}\big(q(t_n)\big)\boldsymbol{\Delta}_h\mathbf{e}_n^q + \mathbf{Z}(q(t_n))\big(\mathbf{e}_n^q, \mathbf{v}(t_n)\big)\Big) - \\ -\mathbf{B}\big(q(t_n)\big)\big((0.5-\beta)\mathbf{e}_n^{\mathbf{a}} - \beta\mathbf{e}_{n+1}^{\mathbf{a}}\big) \tag{85}$$

and a vector-valued function

$$\mathbf{r}_h(t_n, \mathbf{e}_n^q) := \frac{1}{h}\Big(\mathbf{Z}(q(t_n))\big(\mathbf{e}_n^q, \mathbf{v}(t_n)\big) + \widehat{\mathbf{e}}_n^q\mathbf{v}(t_n)\Big) \tag{86}$$

that is linear in $\mathbf{e}_n^q$. Here, the term $\widehat{\mathbf{e}}_n^q\mathbf{v}(t_n) \in \mathbb{R}^k$ represents the matrix commutator $[\widetilde{\mathbf{e}}_n^q, \widetilde{\mathbf{v}}(t_n)] \in \mathfrak{g}$, see (29). By purpose, the notation $\mathbf{r}_n^{\mathbf{B}}$ in (84) adopts the notation r_n that was introduced in (48) to denote a scaled linear combination of global errors in v and local errors in q for proving second-order convergence for the linear test equation, see Sect. 3.2.

The definitions of $\mathbf{r}_n^{\mathbf{B}}$ and $\mathbf{r}_h(t_n, \mathbf{e}_n^q)$ contain a term $\mathbf{Z}(q(t_n))\big(\mathbf{e}_n^q, \mathbf{v}(t_n)\big)/h$ with the bilinear form $\mathbf{Z}(q)$ that is known from the hidden constraints (18) at the level of acceleration coordinates. A time discrete approximation of these hidden constraints shows that the first term in the right-hand side of (85) is of size $\mathcal{O}(1)(\|\mathbf{e}_n^q\| + \|\boldsymbol{\Delta}_h\mathbf{e}_n^q\|)$, see (88):

Lemma 4.7 (Arnold et al. 2015, Lemma 4) *The global errors* $\mathbf{e}_n^q \in \mathbb{R}^k$ *satisfy*

$$\mathbf{B}\big(q(t_n)\big)\mathbf{e}_n^q = \mathcal{O}(h)\|\mathbf{e}_n^q\|\,, \tag{87}$$

$$\mathbf{B}\big(q(t_n)\big)\boldsymbol{\Delta}_h\mathbf{e}_n^q + \mathbf{Z}\big(q(t_n)\big)\big(\mathbf{e}_n^q, \mathbf{v}(t_n)\big) = \mathcal{O}(h)(\|\mathbf{e}_n^q\| + \|\boldsymbol{\Delta}_h\mathbf{e}_n^q\|)\,. \tag{88}$$

Proof Taking into account that $\boldsymbol{\Phi}(q(t_n)) = \boldsymbol{\Phi}(q_n) = \mathbf{0}$, we consider $\boldsymbol{\Phi}(q_{n,\vartheta})$ for $q_{n,\vartheta} := q(t_n) \circ \exp(-\vartheta\widetilde{\mathbf{e}}_n^q) \in G$, ($\vartheta \in [0, 1]$), and get

$$\mathbf{0} = -\big(\boldsymbol{\Phi}(q_n) - \boldsymbol{\Phi}(q(t_n))\big) = -\big(\boldsymbol{\Phi}(q_{n,1}) - \boldsymbol{\Phi}(q_{n,0})\big) = \int_0^1 \mathbf{B}(q_{n,\vartheta})\mathbf{e}_n^q \,\mathrm{d}\vartheta \tag{89}$$

since $\mathbf{B}(q_{n,\vartheta})\mathbf{e}_n^q = -(\mathrm{d}/\mathrm{d}\vartheta)\boldsymbol{\Phi}(q_{n,\vartheta})$, see (14). Assertion (87) follows from (89) because $\mathbf{B}(q_{n,\vartheta}) = \mathbf{B}\big(q(t_n)\big) + \mathcal{O}(h)$, see (77).

The proof of (88) is technically much more complicated and starts with the observation that

$$\mathbf{0} = \int_0^1 \frac{\mathbf{B}(q_{n+1,\vartheta})\mathbf{e}_{n+1}^q - \mathbf{B}(q_{n,\vartheta})\mathbf{e}_n^q}{h}\,\mathrm{d}\vartheta\,,$$

see (89). The integrand may be split into terms $\mathbf{B}(q_{n+1,\vartheta})(\mathbf{e}_{n+1}^q - \mathbf{e}_n^q)/h$ and $\big(\mathbf{B}(q_{n+1,\vartheta})\mathbf{e}_n^q - \mathbf{B}(q_{n,\vartheta})\mathbf{e}_n^q\big)/h$ that yield in (88) the terms $\mathbf{B}\big(q(t_n)\big)\boldsymbol{\Delta}_h\mathbf{e}_n^q$ and $\mathbf{Z}\big(q(t_n)\big)$ $\big(\mathbf{e}_n^q, \mathbf{v}(t_n)\big)$, respectively. For the detailed proof, we refer to (Arnold et al. 2015). □

Lemma 4.8 (Arnold et al. 2015, Lemma 6) *If* $\alpha_m \neq 1$, $\alpha_f \neq 1$, $\beta \neq 0$ *and the order condition (41) is satisfied then*

$$\mathbf{r}_n^{\mathbf{B}} + (0.5-\beta)\mathbf{e}_n^{\mathbf{Ba}} + \beta\mathbf{e}_{n+1}^{\mathbf{Ba}} = \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2)\,, \tag{90}$$

$$(1-\gamma)\mathbf{e}_n^{\mathbf{Ba}} + \gamma\mathbf{e}_{n+1}^{\mathbf{Ba}} = \mathbf{r}_{n+1}^{\mathbf{B}} - \mathbf{r}_n^{\mathbf{B}} + \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2)\,. \tag{91}$$

Proof (a) Inserting error estimate (88) in (85), we get

$$\mathbf{r}_n^{\mathbf{B}} + (0.5-\beta)\mathbf{e}_n^{\mathbf{Ba}} + \beta\mathbf{e}_{n+1}^{\mathbf{Ba}} = \mathcal{O}(1)(\|\mathbf{e}_n^q\| + \|\Delta_h\mathbf{e}_n^q\| + h\|\mathbf{e}_{n+1}^{\mathbf{a}}\|)\,,$$

and (90) follows from (76b).

(b) With the assumptions on the algorithmic parameters α_m, α_f, β and γ, we may substitute in (84) the term $\mathcal{O}(h)\|\mathbf{e}_{n+1}^{\mathbf{a}}\|$ by its upper bound $\mathcal{O}(1)\varepsilon_n + \mathcal{O}(h^2)$, see (Arnold et al. 2015, Corollary 1a). In this modified form, estimate (84) implies

$$\frac{\mathbf{e}_{n+1}^{\mathbf{Bv}} - \mathbf{e}_n^{\mathbf{Bv}}}{h} = \mathbf{r}_{n+1}^{\mathbf{B}} - \mathbf{r}_n^{\mathbf{B}} + \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2) \tag{92}$$

since $\|\mathbf{l}_{n+1}^q - \mathbf{l}_n^q\| = \mathcal{O}(h^4)$, see Lemma 4.2, and $\mathbf{r}_h(t_n, \mathbf{e}_n^q) = (\ldots)/h$ varies smoothly in n in the sense that

$$\begin{aligned}
&\mathbf{r}_h(t_{n+1}, \mathbf{e}_{n+1}^q) - \mathbf{r}_h(t_n, \mathbf{e}_n^q)\\
&= \big(\mathbf{r}_h(t_{n+1}, \mathbf{e}_{n+1}^q) - \mathbf{r}_h(t_{n+1}, \mathbf{e}_n^q)\big) + \big(\mathbf{r}_h(t_{n+1}, \mathbf{e}_n^q) - \mathbf{r}_h(t_n, \mathbf{e}_n^q)\big)\\
&= h\mathbf{r}_h(t_{n+1}, \Delta_h\mathbf{e}_n^q) + h\dot{\mathbf{r}}_h(t_n + \vartheta h, \mathbf{e}_n^q) = \mathcal{O}(1)\|\Delta_h\mathbf{e}_n^q\| + \mathcal{O}(1)\|\mathbf{e}_n^q\|
\end{aligned}$$

with some $\vartheta \in (0,1)$, see also the more detailed discussion in (Arnold et al. 2015, Lemma 6). Inserting (92) into (83), we get estimate (91). □

Finally, a one-step error recursion for the generalized-α Lie group integrator (37) may be formulated in terms of $\mathbf{r}_n^{\mathbf{B}}$ and the vector-valued global errors $\mathbf{e}_n^q$, $\mathbf{e}_n^{\mathbf{v}}$, $\mathbf{e}_n^{\mathbf{Pa}}$, $\mathbf{e}_n^{\mathbf{Ba}}$, $\mathbf{e}_n^{\mathbf{S}\boldsymbol{\lambda}}$ combining (71a), (76a), (81), (82), (90) and (91) to

$$\|\mathbf{E}_{n+1}^{\mathbf{y}} - \mathbf{T}_{\mathbf{y}}\mathbf{E}_n^{\mathbf{y}}\| \le \mathcal{O}(h)(\varepsilon_n + \varepsilon_{n+1} + \|\mathbf{E}_n^{\mathbf{z}}\| + \|\mathbf{E}_{n+1}^{\mathbf{z}}\|) + \mathcal{O}(h^3)\,, \tag{93a}$$

$$\|\mathbf{E}_{n+1}^{\mathbf{z}} - \mathbf{T}_{\mathbf{z}}\mathbf{E}_n^{\mathbf{z}}\| \le \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2) \tag{93b}$$

with

$$\mathbf{E}_n^{\mathbf{y}} := \begin{pmatrix} \mathbf{e}_n^q \\ \mathbf{e}_n^{\mathbf{v}} \end{pmatrix}, \quad \mathbf{E}_n^{\mathbf{z}} := \begin{pmatrix} \mathbf{e}_n^{\mathbf{Pa}} \\ \mathbf{E}_n^{\mathbf{r}} \end{pmatrix}, \quad \mathbf{E}_n^{\mathbf{r}} := \begin{pmatrix} \mathbf{e}_n^{\mathbf{S}\boldsymbol{\lambda}} \\ \mathbf{r}_n^{\mathbf{B}} \\ \mathbf{e}_n^{\mathbf{Ba}} \end{pmatrix}, \tag{94}$$

$$\mathbf{T}_{\mathbf{y}} := \mathbf{I}_{2k}\,, \quad \mathbf{T}_{\mathbf{z}} := \text{blockdiag}\,(-\frac{\alpha_m}{1-\alpha_m})\mathbf{I}_k,\ (\mathbf{T}_+^{-1}\mathbf{T}_0 \otimes \mathbf{I}_m)\,) \tag{95}$$

and

$$\mathbf{T}_+ := \begin{pmatrix} 0 & 0 & -\beta \\ 0 & 1 & -\gamma \\ 1-\alpha_f & 0 & 1-\alpha_m \end{pmatrix}, \quad \mathbf{T}_0 := \begin{pmatrix} 0 & 1 & 0.5-\beta \\ 0 & 1 & 1-\gamma \\ -\alpha_f & 0 & -\alpha_m \end{pmatrix}.$$

The one-step error recursion (93) couples the convergence analysis for unconstrained systems (error components $\mathbf{e}_n^q$, $\mathbf{e}_n^{\mathbf{v}}$, $\mathbf{e}_n^{\mathbf{Pa}}$) to error bounds for the Lagrange multipliers and other algebraic variables (error components $\mathbf{e}_n^{\boldsymbol{\lambda}}$, $\mathbf{r}_n^{\mathbf{B}}$, $\mathbf{e}_n^{\mathbf{Ba}}$). The latter ones are closely related to the error analysis for the linear test equation $\ddot{q} + \omega^2 q = 0$ in the limit case $h\omega \to \infty$, see Eqs. (46)–(48) in Sect. 3.2.

The error bounds (93) are the key to the convergence analysis of the DAE Lie group integrator (37), see Sect. 4.2 and Theorem 4.18 below. In the following, we will call this integrator the *index-3 integrator* since it results from the direct time discretization of the original index-3 formulation (15) of the equations of motion. With a slightly different definition of vectors $\mathbf{E}_n^{\mathbf{r}}$ and matrix $\mathbf{T}_{\mathbf{z}}$, error bounds (93) may also be proved for the *stabilized index-2 integrator* (56) that is based on the stabilized index-2 formulation (55) of the equations of motion. For this integrator, the time discrete approximation of hidden constraints yields:

Lemma 4.9 (see Arnold et al. 2015, Theorem 2)

(a) The auxiliary variables $\boldsymbol{\eta}_n$ in (56b) are of size $\|\boldsymbol{\eta}_n\| = \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2)$. Therefore, error estimate (76a) applies as well to integrator (56).

(b) For integrator (56), the error bounds in (84) and (91) get the form

$$\frac{1}{h}\mathbf{e}_n^{\mathbf{Bv}} = -\bar{\mathbf{r}}_h(t_n, \mathbf{e}_n^q) + \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2)\,, \tag{96}$$

$$(1-\gamma)\mathbf{e}_n^{\mathbf{Ba}} + \gamma\mathbf{e}_{n+1}^{\mathbf{Ba}} = \mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2) \tag{97}$$

with

$$\bar{\mathbf{r}}_h(t_n, \mathbf{e}_n^q) := \frac{1}{h}\mathbf{Z}(q(t_n))\big(\mathbf{v}(t_n), \mathbf{e}_n^q\big)\,.$$

Proof We sketch the basic ideas of the proof and refer to the proof of (Arnold et al. 2015, Theorem 2) for a more detailed discussion.

(a) For the stabilized index-2 formulation, the scaled increment $\Delta_h\mathbf{e}_n^q$ in (73) and (88) has to be substituted by $\Delta_h\mathbf{e}_n^q + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n$, see (56b). In this modified form, estimate (88) yields

$$\mathbf{B}\big(q(t_n)\big)\mathbf{B}^\top(q_n)\boldsymbol{\eta}_n = \mathcal{O}(1)(\|\mathbf{e}_n^q\| + \|\Delta_h\mathbf{e}_n^q\|) \tag{98}$$

with a right-hand side that is of size $\mathcal{O}(1)(\varepsilon_n + \varepsilon_{n+1}) + \mathcal{O}(h^2)$, see (76b). The assertion may be proved solving (98) w.r.t. $\boldsymbol{\eta}_n$ since the full rank assumption on $\mathbf{B}(q)$ implies that $\mathbf{B}\big(q(t_n)\big)\mathbf{B}^\top(q_n) = [\mathbf{B}\mathbf{B}^\top](q_n) + \mathcal{O}(h)$ is non-singular. Using this upper bound for $\|\boldsymbol{\eta}_n\|$, we get error estimate (76a) from $\mathbf{e}_{n+1}^q = \mathbf{e}_n^q + h(\Delta_h\mathbf{e}_n^q + \mathbf{B}^\top(q_n)\boldsymbol{\eta}_n)$.

(b) For the stabilized index-2 formulation, analytical and numerical solution satisfy the hidden constraints (16) resulting in

$$\mathbf{0} = \frac{\mathbf{B}\big(q(t_n)\big)\mathbf{v}(t_n) - \mathbf{B}(q_n)\mathbf{v}_n}{h} = \frac{1}{h}\mathbf{B}(q_n)\mathbf{e}_n^{\mathbf{v}} + \frac{\mathbf{B}\big(q(t_n)\big) - \mathbf{B}(q_n)}{h}\mathbf{v}(t_n) \qquad (99)$$

with $\mathbf{B}(q_n)\mathbf{e}_n^{\mathbf{v}} = \mathbf{e}_n^{\mathbf{Bv}} + \mathcal{O}(h)\varepsilon_n$. For the analysis of the second term in the right-hand side of (99), we use ideas of the proof of Lemma 4.7 and take into account that

$$\big(\mathbf{B}\big(q(t_n)\big) - \mathbf{B}(q_n)\big)\mathbf{v}(t_n) = -\big(\mathbf{B}(q_{n,1}) - \mathbf{B}(q_{n,0})\big)\mathbf{v}(t_n)$$

with $q_{n,\vartheta} := q(t_n) \circ \exp(-\vartheta\widetilde{\mathbf{e}}_n^q)$. Because of

$$-\frac{\mathrm{d}}{\mathrm{d}\vartheta}\big(\mathbf{B}(q_{n,\vartheta})\mathbf{v}(t_n)\big) = \mathbf{Z}(q_{n,\vartheta})\big(\mathbf{v}(t_n), \mathbf{e}_n^q\big) = h\bar{\mathbf{r}}_h(t_n, \mathbf{e}_n^q) + \mathcal{O}(h^2)\varepsilon_n\,,$$

we get $\bar{\mathbf{r}}_h(t_n, \mathbf{e}_n^q) = \big(\mathbf{B}\big(q(t_n)\big) - \mathbf{B}(q_n)\big)\mathbf{v}(t_n)/h + \mathcal{O}(h)\varepsilon_n$ and estimate (96) is seen to be a consequence of (99). With (96), the one-step recursion (97) for error vectors $\mathbf{e}_n^{\mathbf{Ba}}$ may be proved as in Lemma 4.8. □

Because of Lemma 4.9(b), there is no need to consider vectors $\mathbf{r}_n^{\mathbf{B}}$ in the global error analysis of the stabilized index-2 integrator (56). Summarizing error estimates (71a), (76a), (81), (82) and (97), we get the one-step error recursion (93) with

$$\mathbf{T}_{\mathbf{y}} := \mathbf{I}_{2k}\,,\quad \mathbf{T}_{\mathbf{z}} := \mathrm{blockdiag}\,\big(-\frac{\alpha_m}{1-\alpha_m}\mathbf{I}_k,\ (\bar{\mathbf{T}}_+^{-1}\bar{\mathbf{T}}_0 \otimes \mathbf{I}_m)\big) \qquad (100)$$

and

$$\mathbf{E}_n^{\mathbf{r}} := \begin{pmatrix} \mathbf{e}_n^{\mathbf{S}\boldsymbol{\lambda}} \\ \mathbf{e}_n^{\mathbf{Ba}} \end{pmatrix},\quad \bar{\mathbf{T}}_+ := \begin{pmatrix} 0 & -\gamma \\ 1-\alpha_f & 1-\alpha_m \end{pmatrix},\quad \bar{\mathbf{T}}_0 := \begin{pmatrix} 0 & 1-\gamma \\ -\alpha_f & -\alpha_m \end{pmatrix}.$$

4.2 *Coupled Error Propagation in Differential and Algebraic Solution Components*

The classical convergence analysis of ODE one-step methods provides the basis for investigating the coupled error propagation in differential and algebraic solution components of DAE Lie group integrators. We start this section with a perturbation analysis for ODE initial value problems (Theorem 4.10) and consider in Theorem 4.11 the corresponding convergence result for ODE one-step methods. The main new result of this section is the extension of this convergence analysis to the DAE case, see Theorem 4.16.

Theorem 4.10 (see Walter 1998) *Consider the initial value problem*

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t)) \,, \;\; (\, t \in [t_0, t_{\mathrm{end}}]\,) \,, \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{101}$$

with a continuous right-hand side $\mathbf{f}$ *that satisfies for all* $t \in [t_0, t_{\mathrm{end}}]$ *a Lipschitz condition w.r.t.* $\mathbf{x}$ *with a Lipschitz constant* $L > 0$. *For functions* $\hat{\mathbf{x}} \in C^1[t_0, t_{\mathrm{end}}]$ *with*

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{f}(t, \hat{\mathbf{x}}(t)) + \boldsymbol{\delta}(t) \,, \;\; (\, t \in [t_0, t_{\mathrm{end}}]\,) \,, \tag{102}$$

the influence of perturbations $\boldsymbol{\delta}(t)$ *may be estimated by*

$$\|\hat{\mathbf{x}}(t) - \mathbf{x}(t)\| \le \mathrm{e}^{L(t-t_0)} \|\hat{\mathbf{x}}(t_0) - \mathbf{x}(t_0)\| + \frac{\mathrm{e}^{L(t-t_0)} - 1}{L} \max_{s \in [t_0, t_{\mathrm{end}}]} \|\boldsymbol{\delta}(s)\| \,. \tag{103}$$

Proof For $t \in [t_0, t_{\mathrm{end}}]$, we have

$$\begin{aligned} \hat{\mathbf{x}}(t) - \mathbf{x}(t) &= \hat{\mathbf{x}}(t_0) - \mathbf{x}(t_0) + \int_{t_0}^{t} \big(\dot{\hat{\mathbf{x}}}(s) - \dot{\mathbf{x}}(s) \big) \, \mathrm{d}s \\ &= \hat{\mathbf{x}}(t_0) - \mathbf{x}(t_0) + \int_{t_0}^{t} \big(\mathbf{f}(s, \hat{\mathbf{x}}(s)) - \mathbf{f}(s, \mathbf{x}(s)) \big) \, \mathrm{d}s + \int_{t_0}^{t} \boldsymbol{\delta}(s) \, \mathrm{d}s \,. \end{aligned}$$

Therefore, the triangle inequality and the Lipschitz condition on $\mathbf{f}$ imply

$$\|\hat{\mathbf{x}}(t) - \mathbf{x}(t)\| \le \psi(t) \tag{104}$$

with the continuously differentiable function

$$\psi(t) := \|\hat{\mathbf{x}}(t_0) - \mathbf{x}(t_0)\| + L \int_{t_0}^{t} \|\hat{\mathbf{x}}(s) - \mathbf{x}(s)\| \, \mathrm{d}s + (t - t_0)\Delta$$

and $\Delta := \max_{s \in [t_0, t_{\mathrm{end}}]} \|\boldsymbol{\delta}(s)\|$. Note, that $\max_s \|\boldsymbol{\delta}(s)\|$ is well defined since $\hat{\mathbf{x}} \in C^1[t_0, t_{\mathrm{end}}]$ implies that $\boldsymbol{\delta}$ is continuous on the compact interval $[t_0, t_{\mathrm{end}}]$.

Because of (104), the time derivative of ψ satisfies for all $t \in [t_0, t_{\mathrm{end}}]$ the estimate

$$\dot{\psi}(t) = L \|\hat{\mathbf{x}}(t) - \mathbf{x}(t)\| + \Delta \le L\psi(t) + \Delta \,.$$

Hence, the derivative of $\sigma(\tau) := \mathrm{e}^{L(t-\tau)} \psi(\tau)$ is bounded by

$$\sigma'(\tau) = \mathrm{e}^{L(t-\tau)} \big(-L\psi(\tau) + \dot{\psi}(\tau) \big) \le \mathrm{e}^{L(t-\tau)} \, \Delta$$

and we get

$$\sigma(t) = \sigma(s) + \int_{s}^{t} \sigma'(\tau) \, \mathrm{d}\tau \le \sigma(s) + \int_{s}^{t} \mathrm{e}^{L(t-\tau)} \, \mathrm{d}\tau \cdot \Delta \,,$$

i.e.,

$$\psi(t) \le \mathrm{e}^{L(t-s)}\psi(s) + \int_s^t \mathrm{e}^{L(t-\tau)}\,\mathrm{d}\tau \cdot \Delta \tag{105}$$

for any $s \in [t_0, t_{\text{end}}]$. Error bound (105) with $s = t_0$ proves (103) since

$$\int_{t_0}^t \mathrm{e}^{L(t-\tau)}\,\mathrm{d}\tau = \frac{\mathrm{e}^{L(t-t_0)} - 1}{L} \tag{106}$$

and $\psi(t_0) = \|\hat{\mathbf{x}}(t_0) - \mathbf{x}(t_0)\|$. □

For the numerical solution of ODE (101), we consider a one-step method that updates the numerical solution in time step $t_n \to t_{n+1} = t_n + h_n$ according to

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h_n \mathbf{\Phi}_n(t_n, \mathbf{x}_n; \mathbf{f}, h_n) \tag{107}$$

with a continuous increment function $\mathbf{\Phi}$ that satisfies a Lipschitz condition w.r.t. $\mathbf{x}_n$ with a Lipschitz constant $L_{\mathbf{\Phi}} > 0$, see, e.g., (Hairer et al. 1993). The time discretization error in one single time step defines the *local error*

$$\mathbf{le}_n := \mathbf{x}(t_{n+1}) - \big(\mathbf{x}(t_n) + h_n \mathbf{\Phi}(t_n, \mathbf{x}(t_n); \mathbf{f}, h_n)\big)\,.$$

In the global error analysis, the accumulation of these local errors during time integration is studied by a discrete counterpart to the perturbation analysis for the continuous problem (see Theorem 4.10).

Theorem 4.11 *The global errors* $\mathbf{e}_n := \mathbf{x}(t_n) - \mathbf{x}_n$ *satisfy the error recursion*

$$\|\mathbf{e}_{n+1} - \mathbf{e}_n\| \le L_{\mathbf{\Phi}} h_n \|\mathbf{e}_n\| + \|\mathbf{le}_n\| \tag{108}$$

that results in the global error estimate

$$\|\mathbf{e}_n\| \le \mathrm{e}^{L_{\mathbf{\Phi}}(t_n - t_0)}\|\mathbf{e}_0\| + \frac{\mathrm{e}^{L_{\mathbf{\Phi}}(t_n-t_0)} - 1}{L_{\mathbf{\Phi}}} \max_{0 \le l < n} \frac{1}{h_l}\|\mathbf{le}_l\|\,. \tag{109}$$

Proof (a) Using the definition of local and global errors, we get

$$\begin{aligned}
\mathbf{e}_{n+1} - \mathbf{e}_n &= \big(\mathbf{x}(t_{n+1}) - \mathbf{x}_{n+1}\big) - \big(\mathbf{x}(t_n) - \mathbf{x}_n\big) \\
&= \mathbf{x}(t_{n+1}) - \big(\mathbf{x}(t_n) + h_n \mathbf{\Phi}(t_n, \mathbf{x}(t_n); \mathbf{f}, h_n)\big) + \\
&\qquad\qquad + h_n \mathbf{\Phi}(t_n, \mathbf{x}(t_n); \mathbf{f}, h_n) - \big(\mathbf{x}_{n+1} - \mathbf{x}_n\big) \\
&= \mathbf{le}_n + h_n\big(\mathbf{\Phi}(t_n, \mathbf{x}(t_n); \mathbf{f}, h_n) - \mathbf{\Phi}(t_n, \mathbf{x}_n; \mathbf{f}, h_n)\big)\,.
\end{aligned}$$

Therefore, estimate (108) follows from the triangle inequality and from the Lipschitz condition on $\mathbf{\Phi}$:

$$\|\mathbf{e}_{n+1} - \mathbf{e}_n\| \le \|\mathbf{le}_n\| + h_n L_{\mathbf{\Phi}} \|\mathbf{x}(t_n) - \mathbf{x}_n\| = L_{\mathbf{\Phi}} h_n \|\mathbf{e}_n\| + \|\mathbf{le}_n\| \,.$$

(b) Estimate (108) with n being substituted by some $r \in \{0, 1, \dots, n\}$ implies

$$\|\mathbf{e}_{r+1}\| \le \|\mathbf{e}_r\| + L_{\mathbf{\Phi}} h_r \|\mathbf{e}_r\| + \|\mathbf{le}_r\| = (1 + L_{\mathbf{\Phi}} h_r)\|\mathbf{e}_r\| + \|\mathbf{le}_r\| \tag{110}$$

with $h_r = t_{r+1} - t_r$. For a recursive application of this error estimate, we substitute the coefficients of $\|\mathbf{e}_r\|$ and $\|\mathbf{le}_r\|$ in the right-hand side of (110) by upper bounds that are obtained from $1 + Lt \le \mathrm{e}^{Lt}$ and

$$1 = \frac{t_{r+1} - t_r}{h_r} = \frac{1}{h_r} \int_{t_r}^{t_{r+1}} \mathrm{d}\tau \le \frac{1}{h_r} \int_{t_r}^{t_{r+1}} \mathrm{e}^{L_{\mathbf{\Phi}}(t_{r+1} - \tau)} \,\mathrm{d}\tau$$

and get

$$\|\mathbf{e}_{r+1}\| \le \mathrm{e}^{L_{\mathbf{\Phi}}(t_{r+1} - t_r)} \|\mathbf{e}_r\| + \int_{t_r}^{t_{r+1}} \mathrm{e}^{L_{\mathbf{\Phi}}(t_{r+1} - \tau)} \,\mathrm{d}\tau \cdot \frac{1}{h_r} \|\mathbf{le}_r\| \,. \tag{111}$$

(c) Estimate (111) is a special case of the more general expression

$$\|\mathbf{e}_n\| \le \mathrm{e}^{L_{\mathbf{\Phi}}(t_n - t_r)} \|\mathbf{e}_r\| + \int_{t_r}^{t_n} \mathrm{e}^{L_{\mathbf{\Phi}}(t_n - \tau)} \,\mathrm{d}\tau \cdot \max_{r \le l < n} \frac{1}{h_l} \|\mathbf{le}_l\| \,, \tag{112}$$

($r = 0, 1, \dots, n - 1$), that may be considered as a time discrete counterpart to (105). To prove the error bound (112) by induction, we observe that (111) is estimate (112) with $r = n - 1$. For the induction step, we suppose that (112) is satisfied for $r + 1$:

$$\|\mathbf{e}_n\| \le \mathrm{e}^{L_{\mathbf{\Phi}}(t_n - t_{r+1})} \|\mathbf{e}_{r+1}\| + \int_{t_{r+1}}^{t_n} \mathrm{e}^{L_{\mathbf{\Phi}}(t_n - \tau)} \,\mathrm{d}\tau \cdot \max_{r+1 \le l < n} \frac{1}{h_l} \|\mathbf{le}_l\| \,.$$

Inserting in this expression the upper bound (111) for $\|\mathbf{e}_{r+1}\|$, we get estimate (112) since

$$\mathrm{e}^{L_{\mathbf{\Phi}}(t_n - t_{r+1})} \mathrm{e}^{L_{\mathbf{\Phi}}(t_{r+1} - \tau)} = \mathrm{e}^{L_{\mathbf{\Phi}}(t_n - \tau)}$$

for any $\tau \in [t_r, t_{r+1}]$.

(d) To complete the proof, we use the identity (106) and see that (112) with $r = 0$ proves the global error bound (109). □

Abstracting from the specific setting in Theorem 4.11, we may consider more general one-step error recursions and the resulting error bounds. For simplicity, we restrict this analysis to constant time step sizes h. In that case, we may substitute the term $\|\mathbf{le}_r\|$ in (110) by hM with an appropriate constant $M \ge 0$ and get a one-step recursion

$$u_{n+1} \leq (1 + Lh)u_n + hM \,, \tag{113}$$

($n \geq 0$), that implies

$$u_n \leq \mathrm{e}^{L(t_n - t_0)} u_0 + \frac{\mathrm{e}^{L(t_n - t_0)} - 1}{L} M \tag{114}$$

with $u_n := \|\mathbf{e}_n\|$, $L := L_{\mathbf{\Phi}} > 0$ and $t_n := t_0 + nh$, see (109). The convergence analysis of Theorem 4.11 may be generalized straightforwardly to more complex error recursions:

Lemma 4.12 *Consider sequences* $(v_n)_{n \geq 0}$, $(w_n)_{n \geq 0}$ *of non-negative numbers that satisfy*

$$v_{n+1} \leq (1 + Lh)v_n + Lh\kappa^n e_0 + hM \,, \tag{115a}$$

$$w_{n+1} \leq (\kappa + Lh)w_n + Lh\kappa^n e_0 + M \tag{115b}$$

with a positive constant L and non-negative constants $\kappa \in [0, 1)$, *M and* e_0. *All these constants are supposed to be independent of* $h > 0$ *and* $n \geq 0$.

Using the notation $t_n := t_0 + nh$, *we get for all* $n \geq 0$ *the estimate*

$$v_n \leq \mathrm{e}^{L(t_n - t_0)} \big(v_0 + h \, \frac{Le_0}{1 - \kappa} \big) + \frac{\mathrm{e}^{L(t_n - t_0)} - 1}{L} M \,. \tag{116a}$$

For the sequence $(w_n)_{n \geq 0}$, *an estimate*

$$w_n \leq (\kappa + Lh)^n w_0 + h \, \frac{Le_0}{1 - \kappa} + \frac{M}{1 - (\kappa + Lh)} \tag{116b}$$

may be shown for all $n \geq 0$ *and all* $h \in (0, h_0]$ *with* $h_0 > 0$ *denoting a constant such that* $\kappa + Lh_0 < 1$.

Proof Following part (b) of the proof of Theorem 4.11, we rewrite the one-step error recursions (115) in a form that is appropriate for recursive application:

$$v_{r+1} \leq \mathrm{e}^{L(t_{r+1} - t_r)} \big(v_r + Lh\kappa^r e_0 \big) + \int_{t_r}^{t_{r+1}} \mathrm{e}^{L(t_{r+1} - \tau)} \, \mathrm{d}\tau \cdot M \,,$$

$$w_{r+1} \leq (\kappa + Lh)w_r + Lh\kappa^r e_0 + M \,.$$

Then, the error bounds

$$v_n \leq \mathrm{e}^{L(t_n - t_r)} \big(v_r + h \sum_{l=r}^{n-1} \kappa^l \cdot Le_0 \big) + \int_{t_r}^{t_n} \mathrm{e}^{L(t_n - \tau)} \, \mathrm{d}\tau \cdot M \,, \tag{117a}$$

$$w_n \leq (\kappa + Lh)^{n-r} w_r + h \sum_{l=r}^{n-1} \kappa^l \cdot Le_0 + \sum_{l=r}^{n-1} (\kappa + Lh)^{n-(l+1)} \cdot M \,, \tag{117b}$$

($r = 0, 1, \dots, n-1$), follow (similar to part (c) of the proof of Theorem 4.11) by induction starting at $r = n-1$. In the induction step, we have to take into account that

$$\mathrm{e}^{L(t_n - t_r)} \kappa^r + \mathrm{e}^{L(t_n - t_{r+1})} \sum_{l=r+1}^{n-1} \kappa^l \le \mathrm{e}^{L(t_n - t_r)} \sum_{l=r}^{n-1} \kappa^l \, .$$

and $(\kappa + Lh)^{n-(r+1)} < 1$ for any $h \in (0, h_0]$. Error bounds (117) with $r = 0$ prove the lemma since $\kappa \in [0, 1)$ and $\kappa + Lh \in [0, 1)$ imply

$$\sum_{l=r}^{n-1} \kappa^l \le \sum_{l=0}^{\infty} \kappa^l = \frac{1}{1-\kappa} \, , \quad \sum_{l=r}^{n-1} (\kappa + Lh)^{n-(l+1)} \le \frac{1}{1-(\kappa + Lh)}$$

and the integral term in (117a) may be evaluated in closed form, see (106). □

Lemma 4.13 *Let* $(\mathbf{E}_n)_{n \ge 0}$ *be a sequence of vectors that satisfy*

$$\|\mathbf{E}_{n+1} - \mathbf{T}\mathbf{E}_n\| \le L_0 (h \|\mathbf{E}_n\| + h \|\mathbf{E}_{n+1}\|) + h M_0 \tag{118}$$

with a matrix $\mathbf{T}$ *and positive constants* L_0, M_0 *that are independent of* $h > 0$ *and* $n \ge 0$. *If there is a norm* $\|.\|_\varrho$ *such that* $\kappa_\varrho := \|\mathbf{T}\|_\varrho \le 1$ *then (118) implies for time step sizes* $h \in (0, h_0]$ *a one-step recursion*

$$\|\mathbf{E}_{n+1} - \mathbf{T}^{n+1}\mathbf{E}_0\|_\varrho \le (\kappa_\varrho + \tilde{L}_0 h) \|\mathbf{E}_n - \mathbf{T}^n \mathbf{E}_0\|_\varrho + \tilde{L}_0 h \kappa_\varrho^n \|\mathbf{E}_0\|_\varrho + h \tilde{M}_0 \tag{119}$$

and error bounds

$$\|\mathbf{E}_n\| \le \|\mathbf{T}^n \mathbf{E}_0\| + C_0 \|\mathbf{E}_n - \mathbf{T}^n \mathbf{E}_0\|_\varrho \, , \tag{120a}$$

$$\|\mathbf{E}_n\| \le C_0 (\|\mathbf{E}_0\|_\varrho + \|\mathbf{E}_n - \mathbf{T}^n \mathbf{E}_0\|_\varrho) \tag{120b}$$

with appropriate constants h_0, $\tilde{L}_0$, $\tilde{M}_0$ *and* C_0 *that are supposed to be positive. They depend on the norm* $\|.\|$ *and on the constants* L_0, M_0 *in (118).*

Proof (a) Since all norms in a finite-dimensional vector space are equivalent, there are positive constants $\underline{c}$, $\overline{c}$ with

$$\underline{c} \|\mathbf{E}\|_\varrho \le \|\mathbf{E}\| \le \overline{c} \|\mathbf{E}\|_\varrho \tag{121}$$

for any vector $\mathbf{E}$. Therefore, estimate (118) implies

$$\|\mathbf{E}_{n+1} - \mathbf{T}\mathbf{E}_n\|_\varrho \le \hat{L}_0 (h \|\mathbf{E}_n\|_\varrho + h \|\mathbf{E}_{n+1}\|_\varrho) + h \hat{M}_0 \tag{122}$$

with $\hat{L}_0 := \overline{c} L_0 / \underline{c}$, $\hat{M}_0 := M_0 / \underline{c}$.

(b) For the proof of estimate (119), we use the triangle inequality and get

$$\|\mathbf{E}_{n+1} - \mathbf{T}^{n+1}\mathbf{E}_0\|_\varrho \leq \|\mathbf{E}_{n+1} - \mathbf{T}\mathbf{E}_n\|_\varrho + \|\mathbf{T}(\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0)\|_\varrho .$$

The term $\|\mathbf{T}(\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0)\|_\varrho$ is bounded by $\kappa_\varrho\|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho$ with $\kappa_\varrho = \|\mathbf{T}\|_\varrho \leq 1$. We obtain

$$\|\mathbf{E}_{n+1} - \mathbf{T}^{n+1}\mathbf{E}_0\|_\varrho \leq \kappa_\varrho\|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho + \|\mathbf{E}_{n+1} - \mathbf{T}\mathbf{E}_n\|_\varrho$$

and may substitute $\|\mathbf{E}_{n+1} - \mathbf{T}\mathbf{E}_n\|_\varrho$ by the upper bound (122) taking into account that

$$\|\mathbf{E}_n\|_\varrho \leq \|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho + \|\mathbf{T}\|_\varrho^n\,\|\mathbf{E}_0\| = \|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho + \kappa_\varrho^n\|\mathbf{E}_0\|_\varrho .$$

The resulting inequality

$$\begin{aligned}&(1-\hat{L}_0h)\|\mathbf{E}_{n+1} - \mathbf{T}^{n+1}\mathbf{E}_0\|_\varrho \\ &\qquad \leq (\kappa_\varrho + \hat{L}_0h)\|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho + 2\hat{L}_0h\kappa_\varrho^n\|\mathbf{E}_0\|_\varrho + h\hat{M}_0\end{aligned}$$

is multiplied by $1/(1-\hat{L}_0h)$ to get an upper bound for $\|\mathbf{E}_{n+1} - \mathbf{T}^{n+1}\mathbf{E}_0\|_\varrho$. If we suppose that $h \in (0, h_0]$ with $h_0 := 1/(2\hat{L}_0)$ then $1-\hat{L}_0h \geq 1/2$ and we may use the inequalities $(\kappa_\varrho + x)/(1-x) \leq \kappa_\varrho + 4x$ and $1/(1-x) \leq 2$ that are valid for all $x \in [0, 1/2]$. To complete the proof of (119), we set $\tilde{L}_0 := 4\hat{L}_0$ and $\tilde{M}_0 := 2\hat{M}_0$.

(c) Because of $\|\mathbf{E}_n\| \leq \|\mathbf{T}^n\mathbf{E}_0\| + \|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|$, error bound (120a) with $C_0 := \bar{c}$ follows from the equivalence of norms $\|.\|$ and $\|.\|_\varrho$, see (121). With this definition of C_0, we have furthermore $\|\mathbf{E}_n\| \leq C_0\|\mathbf{E}_n\|_\varrho$ and (120b) results from $\|\mathbf{E}_n\|_\varrho \leq \|\mathbf{T}^n\mathbf{E}_0\|_\varrho + \|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho$ with $\|\mathbf{T}^n\|_\varrho \leq \kappa_\varrho^n \leq 1$. □

Corollary 4.14 *If the assumptions of Lemma 4.13 are satisfied with $\kappa_\varrho = \|\mathbf{T}\|_\varrho = 1$ then estimates (119) and (120b) imply*

$$\|\mathbf{E}_n\| \leq \tilde{C}_0\big(\mathrm{e}^{\tilde{L}_0(t_n-t_0)}\|\mathbf{E}_0\| + \frac{\mathrm{e}^{\tilde{L}_0(t_n-t_0)}-1}{\tilde{L}_0}\tilde{M}_0\big) \tag{123}$$

with $t_n := t_0 + nh$, ($n \geq 0$), and a constant $\tilde{C}_0 > 0$ that depends on C_0 and the norm $\|.\|$.

Proof For $\kappa_\varrho = 1$, estimate (119) gets the form (113) with the notations $u_n := \|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho$, $L := \tilde{L}_0$ and $M := \tilde{L}_0\|\mathbf{E}_0\|_\varrho + \tilde{M}_0$. Inserting these expressions in error bound (114), we get

$$\|\mathbf{E}_n - \mathbf{T}^n\mathbf{E}_0\|_\varrho \leq (\mathrm{e}^{\tilde{L}_0(t_n-t_0)} - 1)\|\mathbf{E}_0\|_\varrho + \frac{\mathrm{e}^{\tilde{L}_0(t_n-t_0)}-1}{\tilde{L}_0}\tilde{M}_0$$

since $u_0 = \|\mathbf{E}_0 - \mathbf{T}^0\mathbf{E}_0\|_\varrho = 0$. Therefore, the assertion of the corollary follows directly from (120b) if constant $\tilde{C}_0$ is set to $\tilde{C}_0 := C_0/\min\{1, \underline{c}\}$ such that $C_0\|\mathbf{E}_0\|_\varrho \le \tilde{C}_0\|\mathbf{E}_0\|$ and $C_0\tilde{M}_0 \le \tilde{C}_0\tilde{M}_0$, see (121). □

Remark 4.15 (a) For constant time step sizes $h_n = h =$ const, the convergence result in Theorem 4.11 is a special case of the error analysis in Lemma 4.13 and Corollary 4.14 with $\mathbf{E}_n = \mathbf{e}_n$, $\mathbf{T} = \mathbf{I}$, $\tilde{C}_0 = 1$, $\tilde{L}_0 = L_{\mathbf{\Phi}}$ and $M = \max_l \|\mathbf{le}_l\|/h$.

(b) In ODE time integration, the error estimate of Corollary 4.14 is used to prove the convergence of linear multi-step methods by an equivalent one-step formulation, see (Hairer et al. 1993, Sect. III.4). For a k-step method, vector $\mathbf{E}_n$ is composed of global errors $\mathbf{e}_{n-j}$ at k consecutive grid points $t_{n-(k-1)}, \ldots, t_{n-1}, t_n$ and matrix $\mathbf{T}$ has a Kronecker product structure $\mathbf{T} = \mathbf{A} \otimes \mathbf{I}$ with a companion matrix $\mathbf{A} \in \mathbb{R}^{k\times k}$ that satisfies $\|\mathbf{A}\|_\varrho = 1$ in a suitable norm $\|.\|_\varrho$ if the method is zero-stable. For a more detailed discussion of this convergence analysis, the interested reader is referred to the above cited reference.

(c) For matrices $\mathbf{T}$ with spectral radius $\varrho(\mathbf{T}) = 1$, the transformation to Jordan canonical form may be used to construct a norm $\|.\|_\varrho$ with $\|\mathbf{T}\|_\varrho = 1$ provided that all Jordan blocks corresponding to eigenvalues $\lambda_i[\mathbf{T}]$ with $|\lambda_i[\mathbf{T}]| = 1$ are of dimension 1×1, see (Hairer et al. 1993, Lemma III.4.4).

With appropriate matrices $\mathbf{T}$ of norm $\|\mathbf{T}\|_\varrho = 1$, Lemma 4.13 and Corollary 4.14 provide a unified framework for the error analysis of one-step and multi-step methods in ODE time integration. Corollary 4.14 may be generalized to the technically more challenging DAE case that is characterized by a coupled error propagation in differential and algebraic solution components. The error analysis employs two different error propagation matrices satisfying $\|\mathbf{T}_\mathbf{y}\|_{\mathbf{y},\varrho} = 1$ and $\|\mathbf{T}_\mathbf{z}\|_{\mathbf{z},\varrho} < 1$, respectively. It is inspired by the classical convergence analysis of one-step methods for index-1 DAEs in (Deuflhard et al. 1987), see also (Arnold et al. 2015, Lemma 7).

Theorem 4.16 *Let $(\mathbf{E}^\mathbf{y}_n)_{n\ge 0}$ and $(\mathbf{E}^\mathbf{z}_n)_{n\ge 0}$ be sequences of vectors that satisfy*

$$\|\mathbf{E}^\mathbf{y}_{n+1} - \mathbf{T}_\mathbf{y}\mathbf{E}^\mathbf{y}_n\| \le L_0 h(\|\mathbf{E}^\mathbf{y}_n\| + \|\mathbf{E}^\mathbf{y}_{n+1}\| + \|\mathbf{E}^\mathbf{z}_n\| + \|\mathbf{E}^\mathbf{z}_{n+1}\|) + hM_0, \tag{124a}$$
$$\|\mathbf{E}^\mathbf{z}_{n+1} - \mathbf{T}_\mathbf{z}\mathbf{E}^\mathbf{z}_n\| \le L_0(\|\mathbf{E}^\mathbf{y}_n\| + \|\mathbf{E}^\mathbf{y}_{n+1}\| + h\|\mathbf{E}^\mathbf{z}_n\| + h\|\mathbf{E}^\mathbf{z}_{n+1}\|) + M_0 \tag{124b}$$

with matrices $\mathbf{T}_\mathbf{y}$, $\mathbf{T}_\mathbf{z}$ and positive constants L_0, M_0 that are independent of $h > 0$ and $n \ge 0$. If there are norms $\|.\|_{\mathbf{y},\varrho}$, $\|.\|_{\mathbf{z},\varrho}$ such that $\|\mathbf{T}_\mathbf{y}\|_{\mathbf{y},\varrho} = 1$ and $\|\mathbf{T}_\mathbf{z}\|_{\mathbf{z},\varrho} < 1$ then (124) implies for time step sizes $h \in (0, h_0]$ error bounds

$$\|\mathbf{E}^\mathbf{y}_n\| \le \mathrm{e}^{\bar{L}_0(t_n-t_0)}(\|\mathbf{E}^\mathbf{y}_0\| + \bar{C}_0 h\|\mathbf{E}^\mathbf{z}_0\|) + \frac{\mathrm{e}^{\bar{L}_0(t_n-t_0)} - 1}{\bar{L}_0}\bar{M}_0\,, \tag{125a}$$
$$\|\mathbf{E}^\mathbf{z}_n - \mathbf{T}^n_\mathbf{z}\mathbf{E}^\mathbf{z}_0\| \le \bar{C}_0\mathrm{e}^{\bar{L}_0(t_n-t_0)}(\|\mathbf{E}^\mathbf{y}_0\| + h\|\mathbf{E}^\mathbf{z}_0\| + \bar{M}_0) \tag{125b}$$

with $t_n := t_0 + nh$, ($n \ge 0$). The constants h_0, $\bar{C}_0$, $\bar{L}_0$ and $\bar{M}_0$ are supposed to be positive. They depend on constants L_0, M_0 in (124) and may depend furthermore on the vector norms $\|.\| = \|.\|_\mathbf{y}$ and $\|.\| = \|.\|_\mathbf{z}$ for $\mathbf{E}^\mathbf{y}_n$ and $\mathbf{E}^\mathbf{z}_n$.

Proof (a) Using the same arguments as in parts (a) and (c) of the proof of Lemma 4.13, we may verify that the assertion of the Theorem (with appropriate norm dependent constants $\bar{C}_0$, $\bar{L}_0$ and $\bar{M}_0$) is valid for *any* pair of norms $(\|.\|_{\mathbf{y}}, \|.\|_{\mathbf{z}})$ if it is valid for one specific pair $(\|.\|_{\mathbf{y},*}, \|.\|_{\mathbf{z},*})$. To simplify the notation, we will therefore restrict the error analysis to a pair of norms with $\kappa_{\mathbf{y}} := \|\mathbf{T}_{\mathbf{y}}\|_{\mathbf{y}} = 1$ and $\kappa_{\mathbf{z}} := \|\mathbf{T}_{\mathbf{z}}\|_{\mathbf{z}} < 1$ and will furthermore omit the indices $\mathbf{y}$ and $\mathbf{z}$ at the norm symbol $\|.\|$.

(b) Similar to Lemma 4.13 and Corollary 4.14, the coupled error propagation is studied in terms of sequences $(u_n)_{n\geq 0}$, $(w_n)_{n\geq 0}$ with

$$u_n := \|\mathbf{E}_n^{\mathbf{y}} - \mathbf{T}_{\mathbf{y}}^n \mathbf{E}_0^{\mathbf{y}}\|\,, \quad w_n := \|\mathbf{E}_n^{\mathbf{z}} - \mathbf{T}_{\mathbf{z}}^n \mathbf{E}_0^{\mathbf{z}}\|\,. \tag{126}$$

For a one-step error recursion, we look for error bounds like (119) for u_{n+1} and w_{n+1}. As in Lemma 4.13, we get from assumptions (124) the estimates

$$u_{n+1} \leq (1+\tilde{L}_0 h)u_n + \tilde{L}_0 h w_n + \tilde{L}_0 h \kappa_{\mathbf{z}}^n \|\mathbf{E}_0^{\mathbf{z}}\| + h(\tilde{M}_0 + \tilde{L}_0 \|\mathbf{E}_0^{\mathbf{y}}\|)\,, \tag{127a}$$

$$w_{n+1} \leq \tilde{L}_0 u_n + (\kappa_{\mathbf{z}} + \tilde{L}_0 h) w_n + \tilde{L}_0 h \kappa_{\mathbf{z}}^n \|\mathbf{E}_0^{\mathbf{z}}\| + \tilde{M}_0 + \tilde{L}_0 \|\mathbf{E}_0^{\mathbf{y}}\| \tag{127b}$$

with appropriate positive constants $\tilde{L}_0$ and $\tilde{M}_0$. Here, we have taken into account that $\kappa_{\mathbf{y}} = \|\mathbf{T}_{\mathbf{y}}\| = 1$ and $\kappa_{\mathbf{z}} = \|\mathbf{T}_{\mathbf{z}}\| < 1$ and restricted the analysis to $h \in (0, h_0]$ with a sufficiently small constant $h_0 > 0$.

(c) The recursive application of error bounds (127) shows that the coupled error propagation in differential and algebraic solution components may be studied analysing powers of the 2×2 error amplification matrix

$$\mathbf{W}(h) := \begin{pmatrix} 1+\tilde{L}_0 h & \tilde{L}_0 h \\ \tilde{L}_0 & \kappa_{\mathbf{z}} + \tilde{L}_0 h \end{pmatrix},$$

see (Deuflhard et al. 1987, Lemma 2). The eigenvalue analysis for matrix $\mathbf{W}(h)$ yields an eigenvalue $\lambda(h) = \kappa_{\mathbf{z}} + \mathcal{O}(h)$. Because of $\kappa_{\mathbf{z}} < 1$, this eigenvalue satisfies $\lambda(h) < 1$ for all sufficiently small time step sizes $h > 0$. The corresponding eigenvector

$$\boldsymbol{\zeta}(h) := \begin{pmatrix} -L_v h \\ 1 \end{pmatrix} \text{ with } L_v := \frac{\tilde{L}_0}{1+\tilde{L}_0 h - \lambda(h)} = \frac{\tilde{L}_0}{1-\kappa_{\mathbf{z}}} + \mathcal{O}(h) \tag{128}$$

is used to transform $\mathbf{W}(h)$ to lower triangular form: We define the transformation matrix

$$\mathbf{V}(h) := [\,\mathbf{e}_1 \;\; \boldsymbol{\zeta}(h)\,] = \begin{pmatrix} 1 & -L_v h \\ 0 & 1 \end{pmatrix} \text{ with } \mathbf{V}^{-1}(h) = \begin{pmatrix} 1 & L_v h \\ 0 & 1 \end{pmatrix} \tag{129}$$

and observe that the second column vector of $\mathbf{W}(h)\mathbf{V}(h)$ is a multiple of the second column vector of $\mathbf{V}(h)$ since $\mathbf{W}(h)\boldsymbol{\zeta}(h) = \lambda(h)\boldsymbol{\zeta}(h)$. Therefore, the scalar product of the first row vector of $\mathbf{V}^{-1}(h)$ and the second column vector of $\mathbf{W}(h)\mathbf{V}(h)$, i.e., the upper right element of $\mathbf{V}^{-1}(h)\mathbf{W}(h)\mathbf{V}(h)$, vanishes. Straightforward computations yield

$$\mathbf{V}^{-1}(h)\mathbf{W}(h)\mathbf{V}(h) = \begin{pmatrix} 1 + \tilde{L}_0(L_v + 1)h & 0 \\ \tilde{L}_0 & \kappa_{\mathbf{z}} + \tilde{L}_0(1 - L_v)h \end{pmatrix} \tag{130}$$

and

$$v_{n+1} \leq (1 + \tilde{L}_0(L_v + 1)h)v_n + \tag{131a}$$
$$+ \tilde{L}_0(L_v h + 1)h\kappa_{\mathbf{z}}^n \|\mathbf{E}_0^{\mathbf{z}}\| + (L_v + 1)h(\tilde{M}_0 + \tilde{L}_0\|\mathbf{E}_0^{\mathbf{y}}\|)\,,$$
$$w_{n+1} \leq \tilde{L}_0 v_n + (\kappa_{\mathbf{z}} + \tilde{L}_0(1 - L_v)h)w_n + \tag{131b}$$
$$+ \tilde{L}_0 h\kappa_{\mathbf{z}}^n \|\mathbf{E}_0^{\mathbf{z}}\| + \tilde{M}_0 + \tilde{L}_0\|\mathbf{E}_0^{\mathbf{y}}\|$$

with a sequence $(v_n)_{n\geq 0}$ of non-negative numbers v_n that are defined by

$$\begin{pmatrix} v_n \\ w_n \end{pmatrix} = \mathbf{V}^{-1}(h) \begin{pmatrix} u_n \\ w_n \end{pmatrix},$$

see (127), (130) and (131). Note, that all matrix elements of $\mathbf{V}^{-1}(h)$ are non-negative which is an essential assumption for the transformation from (127) to (131).

(d) The right-hand side of (131a) depends nonlinearly on h because $L_v = L_v(h)$. If we substitute L_v for sufficiently small time step sizes $h > 0$ by the upper bound $\tilde{L}_v := 2\tilde{L}_0/(1 - \kappa_{\mathbf{z}})$, see (128), then Lemma 4.12 may be applied with constants $L := \tilde{L}_0(\tilde{L}_v \max\{1, h_0\} + 1)$, $\kappa := \kappa_{\mathbf{z}} < 1$, $e_0 := \|\mathbf{E}_0^{\mathbf{z}}\|$ and $M := (\tilde{L}_v + 1)\tilde{M}_0 + L\|\mathbf{E}_0^{\mathbf{y}}\|$. Inequality (116a) yields the error bound

$$v_n \leq \mathrm{err}_n - \|\mathbf{E}_0^{\mathbf{y}}\| \tag{132}$$

with

$$\mathrm{err}_n := \mathrm{e}^{L(t_n - t_0)}(\|\mathbf{E}_0^{\mathbf{y}}\| + \frac{hL}{1-\kappa}\|\mathbf{E}_0^{\mathbf{z}}\|) + \frac{\mathrm{e}^{L(t_n - t_0)} - 1}{L}(\tilde{L}_v + 1)\tilde{M}_0 \tag{133}$$

because $v_0 = u_0 + L_v h w_0 = 0$, see (126). Inequality (132) proves the global error bound (125a) since $u_n = v_n - L_v h w_n \leq v_n$ and

$$\|\mathbf{E}_n^{\mathbf{y}}\| \leq \|\mathbf{T}_{\mathbf{y}}^n \mathbf{E}_0^{\mathbf{y}}\| + \|\mathbf{E}_n^{\mathbf{y}} - \mathbf{T}_{\mathbf{y}}^n \mathbf{E}_0^{\mathbf{y}}\| \leq \|\mathbf{T}_{\mathbf{y}}\|^n \|\mathbf{E}_0^{\mathbf{y}}\| + u_n \leq \|\mathbf{E}_0^{\mathbf{y}}\| + v_n \leq \mathrm{err}_n\,.$$

For the proof of error bound (125b), we substitute in (131b) the variable v_n by its upper bound (132) and get

$$w_{n+1} \leq (\kappa + Lh)w_n + Lh\kappa^n e_0 + \tilde{M}_0 + \tilde{L}_0\,\mathrm{err}_n$$

since $\tilde{L}_0(1 - L_v) \le \tilde{L}_0 \le L$. For all $r \le n$, the term $\tilde{M}_0 + \tilde{L}_0\,\mathrm{err}_r$ is bounded by $\tilde{M}_0 + \tilde{L}_0\,\mathrm{err}_n$ because $(\mathrm{err}_n)_{n\ge 0}$ is monotonically increasing. Therefore, Lemma 4.12 with

$$M := \tilde{M}_0 + \tilde{L}_0\,\mathrm{err}_n \le \tilde{L}_0 \mathrm{e}^{L(t_n - t_0)}(\|\mathbf{E}_0^{\mathbf{y}}\| + \frac{hL}{1-\kappa}\|\mathbf{E}_0^{\mathbf{z}}\|) + \mathrm{e}^{L(t_n-t_0)}\tilde{M}_0\,,$$

see (133), yields

$$w_n \le C_0^{\mathbf{z}}\big(h\|\mathbf{E}_0^{\mathbf{z}}\| + \mathrm{e}^{L(t_n-t_0)}(\|\mathbf{E}_0^{\mathbf{y}}\| + \frac{hL}{1-\kappa}\|\mathbf{E}_0^{\mathbf{z}}\| + \tilde{M}_0)\big)$$

with an appropriate constant $C_0^{\mathbf{z}} > 0$, see (116b). Error bound (125b) follows straightforwardly from $\|\mathbf{E}_n^{\mathbf{z}} - \mathbf{T}_{\mathbf{z}}^n\mathbf{E}_0^{\mathbf{z}}\| = w_n$, see (126). □

4.3 Convergence of Lie Group Time Integration Methods

For the application of Theorem 4.16 to the one-step error recursion (93) we have to verify the assumptions on error propagation matrices $\mathbf{T}_{\mathbf{y}}$ and $\mathbf{T}_{\mathbf{z}}$. Because of $\mathbf{T}_{\mathbf{y}} = \mathbf{I}_{2k}$, we get $\|\mathbf{T}_{\mathbf{y}}\|_2 = 1$. For proving $\|\mathbf{T}_{\mathbf{z}}\|_{\mathbf{z},\varrho} < 1$ in a suitable norm $\|.\|_{\mathbf{z},\varrho}$, we analyse the spectral radius $\rho(\mathbf{T}_{\mathbf{z}})$:

Lemma 4.17 *(a) For algorithmic parameters α_m, α_f, β, γ that satisfy the order condition (41) and the stability conditions*

$$\alpha_m < \alpha_f < 0.5\,, \qquad \gamma < 2\beta\,, \tag{134}$$

the spectral radii of matrices $\mathbf{T}_{\mathbf{z}}$ in (95) and (100) are bounded by $\rho(\mathbf{T}_{\mathbf{z}}) < 1$.

(b) For the "optimal" parameters of Chung and Hulbert (1993), see (42), the stability conditions (134) are satisfied for any $\rho_\infty \in [0, 1)$.

Proof (a) The block-diagonal structure of matrix $\mathbf{T}_{\mathbf{z}} \in \mathbb{R}^{m+3k}$ in (95) implies that its characteristic polynomial is given by

$$\det(\zeta\mathbf{I} - \mathbf{T}_{\mathbf{z}}) = \Big(\zeta + \frac{\alpha_m}{1-\alpha_m}\Big)^k \Big(\det \mathbf{T}_+^{-1}\det(\zeta\mathbf{T}_+ - \mathbf{T}_0)\Big)^m\,.$$

Straightforward computations show that matrix $\mathbf{T}_{\mathbf{z}}$ has an eigenvalue $\zeta_m := -\alpha_m/(1-\alpha_m)$ of multiplicity k, an eigenvalue $\zeta_f := -\alpha_f/(1-\alpha_f)$ of multiplicity m and eigenvalues $\zeta_{1,2}$ that are given by the roots of the quadratic polynomial $\sigma(\zeta) := a\zeta^2 + b\zeta + c$ with

$$a := \beta\,, \quad b := 0.5 + \gamma - 2\beta\,, \quad c := 1 - a - b\,, \tag{135}$$

see also (Arnold and Brüls 2007, Lemma 1). The stability conditions (134) imply $|\zeta_m| < 1$, $|\zeta_f| < 1$ and $\gamma = 0.5 + \alpha_f - \alpha_m > 0.5$.

Therefore, the coefficients a, b, c in (135) satisfy $a = \beta > 0$, $b > 1 - 2\beta = 1 - 2a$ and $c = 1 - a - b < a$. Since $c/a < 1$ and $\zeta_1\zeta_2 = c/a$ (Vieta's theorem), we get $|\zeta_1|^2 = |\zeta_2|^2 = \zeta_1\zeta_2 = c/a < 1$ whenever $\sigma(\zeta) = 0$ has a pair of conjugate complex roots ζ_1, ζ_2.

If both roots of σ are real then the discriminant

$$b^2 - 4ac = b^2 - 4a(1 - a - b) = (2a + b)^2 - 4a$$

has to be non-negative. Hence,

$$\sqrt{b^2 - 4ac} < \sqrt{(2a + b)^2} = 2a + b \tag{136a}$$

since $a > 0$ and $2a + b = 0.5 + \gamma > 1 \geq 0$, see (135). On the other hand, stability condition $\gamma < 2\beta$ results in $b < 0.5$ and

$$(2a + b)^2 - 4a = (2a - b)^2 + 8a(b - 0.5) < (2a - b)^2 \, ,$$

i.e.,

$$\sqrt{b^2 - 4ac} = \sqrt{(2a + b)^2 - 4a} < \sqrt{(2a - b)^2} = 2a - b \tag{136b}$$

since $2a - b = 2(2\beta - \gamma) + (\gamma - 0.5) > 0$. Estimates (136) show that the roots $\zeta_{1,2} = (-b \pm \sqrt{b^2 - 4ac})/2a$ of σ satisfy $-1 < \zeta_i < 1$, $(\, i = 1, 2\,)$. This completes the proof of $\rho(\mathbf{T_z}) < 1$ for matrix $\mathbf{T_z}$ being defined in (95).

Substituting the quadratic polynomial $\sigma(\zeta)$ by $\sigma(\zeta) := \zeta + (1 - \gamma)/\gamma$, we may extend this analysis straightforwardly to the matrix $\mathbf{T_z}$ in (100).

(b) With $\rho_\infty \in [0, 1)$, the algorithmic parameters α_m, α_f in (42) satisfy $\alpha_m < \alpha_f < 0.5$ and $\gamma = 0.5 + \alpha_f - \alpha_m > 0.5$. For the second stability condition in (134), we observe that (42) implies $2\beta - \gamma = (\gamma - 0.5)^2/2 > 0$. □

Theorem 4.18 *Let the order condition (41) and the stability conditions (134) be fulfilled and suppose that the starting values* q_0, $\mathbf{v}_0$, $\dot{\mathbf{v}}_0$, $\mathbf{a}_0$ *and* $\boldsymbol{\lambda}_0$ *satisfy*

$$\|\mathbf{e}_0^q\| + \|\mathbf{e}_0^{\mathbf{v}}\| + h\|\mathbf{e}_0^{\mathbf{Pa}}\| = \mathcal{O}(h^2)\,, \quad \|\mathbf{e}_0^{\dot{\mathbf{v}}}\| + \|\mathbf{e}_0^{\mathbf{Ba}}\| = \mathcal{O}(h^{1+\delta})\,, \tag{137a}$$

$$\|\mathbf{M}(q_0)\dot{\mathbf{v}}_0 + \mathbf{g}(q_0, \mathbf{v}_0, t_0) + \mathbf{B}^\top(q_0)\boldsymbol{\lambda}_0\| = \mathcal{O}(h^{1+\delta}) \tag{137b}$$

with a non-negative constant $\delta \in [0, 1]$. *Then, there are positive constants* C_0, $\tilde{L}$, h_0 *being independent of* n *and* h *such that we have for all* $h \in (0, h_0]$ *and all* $n \geq 0$ *with* $t_0 + nh \leq t_{\text{end}} - h$*:*

(a) a global error bound

$$\|\mathbf{e}_n^q\| + \|\mathbf{e}_n^{\mathbf{v}}\| \le C_0 \mathrm{e}^{\tilde{L}(t_n-t_0)} h^2 \,, \tag{138a}$$

$$\|\mathbf{e}_n^{\boldsymbol{\lambda}}\| \le C_0 (\|(\mathbf{T}_+^{-1}\mathbf{T}_0)^n\| h^{1+\delta} + \mathrm{e}^{\tilde{L}(t_n-t_0)} h^2) \tag{138b}$$

for the index-3 integrator (37) provided that the starting values q_0, $\mathbf{v}_0$ satisfy the additional assumption

$$\|\mathbf{e}_0^q\| + \|\mathbf{e}_0^{\mathbf{Bv}} + \frac{1}{h}\mathbf{B}(q(t_0))\mathbf{l}_0^q\| = \mathcal{O}(h^{2+\delta}) \tag{139}$$

and

(b) a global error bound

$$\|\mathbf{e}_n^q\| + \|\mathbf{e}_n^{\mathbf{v}}\| + \|\boldsymbol{\eta}_n\| \le C_0 \mathrm{e}^{\tilde{L}(t_n-t_0)} h^2 \,, \tag{140a}$$

$$\|\mathbf{e}_n^{\boldsymbol{\lambda}}\| \le C_0 (\|(\bar{\mathbf{T}}_+^{-1}\bar{\mathbf{T}}_0)^n\| h^{1+\delta} + \mathrm{e}^{\tilde{L}(t_n-t_0)} h^2) \tag{140b}$$

for the stabilized index-2 integrator (56).

Proof These error estimates are a straightforward consequence of Theorem 4.16 and Lemma 4.17 since error recursion (93) with matrices $\mathbf{T}_{\mathbf{y}}$ and $\mathbf{T}_{\mathbf{z}}$ being defined in (95), (100) and $\varepsilon_n = \mathcal{O}(1)(\|\mathbf{E}_n^{\mathbf{y}}\| + h\|\mathbf{E}_n^{\mathbf{z}}\|)$ imply (124). Furthermore, assumptions (137) and (139) result in $\|\mathbf{E}_0^{\mathbf{y}}\| = \mathcal{O}(h^2)$, $\|\mathbf{E}_0^{\mathbf{z}}\| = \mathcal{O}(h)$ and $\|\mathbf{E}_0^{\mathbf{r}}\| = \mathcal{O}(h^{1+\delta})$. Finally, the upper bound for $\|\boldsymbol{\eta}_n\|$ in (140a) is obtained from (98). □

Lemma 4.17 and Theorem 4.18 show that transient errors of size $\mathcal{O}(h^{1+\delta})$ are damped out by numerical dissipation if the generalized-α methods (37) and (56) have algorithmic parameters according to (42) with $\rho_\infty < 1$. For starting values $q_0 = q(t_0)$, $\mathbf{v}_0 = \mathbf{v}(t_0)$, $\dot{\mathbf{v}}_0 = \dot{\mathbf{v}}(t_0)$ and $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}(t_0)$ being defined by consistent initial values $q(t_0)$, $\mathbf{v}(t_0)$, $\dot{\mathbf{v}}(t_0)$, $\boldsymbol{\lambda}(t_0)$, assumptions (137) and (139) are satisfied with $\delta \ge 0$ if $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0) + \mathcal{O}(h)$. Beyond the transient phase, we observe second-order convergence in all solution components, see Fig. 8.

For the heavy top benchmark problem in configuration space $G = \mathrm{SE}(3)$, we may even prove that there is no order reduction at all in generalized-α Lie group time integration:

Example 4.19 (a) For consistent initial values $q(t_0)$, $\mathbf{v}(t_0)$, $\dot{\mathbf{v}}(t_0)$ and $\boldsymbol{\lambda}(t_0)$, the starting values $q_0 = q(t_0)$, $\mathbf{v}_0 = \mathbf{v}(t_0)$, $\dot{\mathbf{v}}_0 = \dot{\mathbf{v}}(t_0)$, $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0)$, $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}(t_0)$ satisfy assumption (137) with $\delta = 1$ if $\mathbf{B}\big(q(t_0)\big)\ddot{\mathbf{v}}(t_0) = \mathbf{0}$ since Taylor expansion of $\dot{\mathbf{v}}(t_0 + \Delta_\alpha h)$ at $h = 0$ shows in that case that $\|\mathbf{e}_0^{\mathbf{Ba}}\| = \|\mathbf{B}\big(q(t_0)\big)\big(\dot{\mathbf{v}}(t_0 + \Delta_\alpha h) - \mathbf{a}_0\big)\| = \mathcal{O}(h^2)$.

(b) Condition $\mathbf{B}\big(q(t_0)\big)\ddot{\mathbf{v}}(t_0) = \mathbf{0}$ in part (a) of this example is satisfied for the equations of motion (22) of the heavy top benchmark in configuration space $G = \mathrm{SE}(3)$ since $\mathbf{B}\big(q(t)\big) \equiv \mathbf{B}^{\mathbf{X}} := \big(-\widetilde{\mathbf{X}} \quad -\mathbf{I}_3\big)$ along any solution curve $q(t)$ in the constraint manifold $\mathfrak{M} := \{\, q \,:\, \boldsymbol{\Phi}(q) = \mathbf{0} \,\}$, see Lemma 3.5, and the hidden constraints (16),

(18) are given by $\mathbf{0} = \mathbf{B}^{\mathbf{X}}\mathbf{v}(t) = \mathbf{B}^{\mathbf{X}}\dot{\mathbf{v}}(t)$ implying $\mathbf{B}\big(q(t)\big)\ddot{\mathbf{v}}(t) = \mathbf{0}$. Therefore, Theorem 4.18(b) proves second-order convergence of the stabilized index-2 integrator (56) for this benchmark problem. These theoretical investigations are illustrated by the numerical test results in the right plot of Fig. 11.

(c) The equations of motion (22) of the heavy top benchmark in configuration space $G = \mathrm{SE}(3)$ fulfill the assumptions of Lemma 3.5. Therefore, the generalized-α integrator (37) defines a numerical solution that satisfies the hidden constraints (16) at the level of velocity coordinates. I.e., integrators (37) and (56) define identical numerical solutions for this benchmark problem and we get $\boldsymbol{\eta}_n = \mathbf{0}$. The numerical test results in the right plots of Figs. 6 and 11 illustrate this coincidence.

For a more direct proof of the corresponding second-order convergence result for integrator (37), we may verify that for this benchmark problem assumption (139) in Theorem 4.18(a) is satisfied with $\delta = 1$: Taking into account $\mathbf{B}\big(q(t_n)\big)\ddot{\mathbf{v}}(t_n) = \mathbf{0}$ and the structure of the leading error term in $\mathbf{l}_n^q$, we get $\mathbf{B}\big(q(t_n)\big)\mathbf{l}_n^q = \mathcal{O}(h^4)$ if $\mathbf{B}\big(q(t)\big)\widehat{\mathbf{v}}(t)\dot{\mathbf{v}}(t) \equiv \mathbf{0}$, see Lemma 4.2. Here, we have substituted the term $[\widetilde{\mathbf{v}}, \widetilde{\dot{\mathbf{v}}}] \in \mathfrak{se}(3)$ in (69) by its equivalent $\widehat{\mathbf{v}}\dot{\mathbf{v}} \in \mathbb{R}^6$ with $\widehat{\mathbf{v}} \in \mathbb{R}^{6\times 6}$ being defined in (34), see also (29). For consistent velocity vectors $\mathbf{v}$, the skew symmetric matrix $\widetilde{\mathbf{U}}$ in (34) may be expressed in terms of $\widetilde{\mathbf{X}}$ and $\widetilde{\boldsymbol{\Omega}}$ since $\mathbf{B}^{\mathbf{X}}\mathbf{v} = \mathbf{0}$ implies $\mathbf{U} = -\widetilde{\mathbf{X}}\boldsymbol{\Omega} = \widetilde{\boldsymbol{\Omega}}\mathbf{X} = \widehat{\boldsymbol{\Omega}}\mathbf{X}$, i.e., $\widetilde{\mathbf{U}} = [\widetilde{\boldsymbol{\Omega}}, \widetilde{\mathbf{X}}] = \widetilde{\boldsymbol{\Omega}}\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\Omega}}$, see (29). The identity $\widetilde{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}$ is valid for any $\boldsymbol{\Omega} \in \mathbb{R}^3$, see Remark 2.8(b). We get

$$\mathbf{B}\big(q(t)\big)\widehat{\mathbf{v}}(t) = \mathbf{B}^{\mathbf{X}}\begin{pmatrix} \widetilde{\boldsymbol{\Omega}} & \mathbf{0} \\ \widetilde{\mathbf{U}} & \widetilde{\boldsymbol{\Omega}} \end{pmatrix} = \big(-\widetilde{\mathbf{X}} \;\; -\mathbf{I}_3\,\big)\begin{pmatrix} \widetilde{\boldsymbol{\Omega}} & \mathbf{0} \\ \widetilde{\boldsymbol{\Omega}}\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\Omega}} & \widetilde{\boldsymbol{\Omega}} \end{pmatrix} = \widetilde{\boldsymbol{\Omega}}\mathbf{B}^{\mathbf{X}}$$

and therefore also $\mathbf{B}\big(q(t)\big)\widehat{\mathbf{v}}(t)\dot{\mathbf{v}}(t) \equiv \mathbf{0}$ since $\mathbf{B}^{\mathbf{X}}\dot{\mathbf{v}}(t) \equiv \mathbf{0}$, see (18). Hence, $\mathbf{B}\big(q(t_n)\big)$ $\mathbf{l}_n^q = \mathcal{O}(h^4)$ and assumptions (139) are satisfied for this benchmark problem with $\delta = 1$ if the starting values in the index-3 integrator (37) are set to $q_0 = q(t_0)$, $\mathbf{v}_0 = \mathbf{v}(t_0)$.

Example 4.19 illustrates that the trivial initialization $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0)$ results for certain problem classes in transient error terms of size $\mathcal{O}(h^{1+\delta})$ with $\delta = 1$ such that second-order convergence is already observed in the transient phase. In general, however, this trivial initialization yields transient errors of size $\mathcal{O}(h)$ since $\|\mathbf{e}_0^{\mathbf{Ba}}\| = \mathcal{O}(h)$ if $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0)$ and $\mathbf{B}\big(q(t_0)\big)\ddot{\mathbf{v}}(t_0) \neq \mathbf{0}$. These first order error terms have been observed numerically for the heavy top benchmark problem in configuration space $G = \mathrm{SO}(3) \times \mathbb{R}^3$ in Figs. 6, 7 and 13.

More sophisticated initializations of sequence $(\mathbf{a}_n)_{n\geq 0}$ in HHT-α and generalized-α time integration have been discussed, e.g., in (Jay and Negrut 2007) and (Arnold et al. 2015). We follow the latter approach and set

$$\mathbf{a}_0 := \dot{\mathbf{v}}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{a}} \quad \text{with} \quad \boldsymbol{\Delta}_0^{\mathbf{a}} := \Delta_\alpha h \,\frac{\dot{\mathbf{v}}_{sh} - \dot{\mathbf{v}}_{-sh}}{2sh}\,, \tag{141}$$

vectors $\dot{\mathbf{v}}_{\pm sh} = \dot{\mathbf{v}}(t_0 \pm sh) + \mathcal{O}(h^2)$ and a (small) parameter $s \in (0, 1]$ that may be set, e.g., to $s := 1/10$. For the computation of $\boldsymbol{\Delta}_0^{\mathbf{a}}$, we have to evaluate the equations

Table 1 Initialization of the stabilized index-2 integrator (56)

Data	Consistent initial values $q(t_0)$, $\mathbf{v}(t_0)$; parameter $s \in (0, 1]$
Result	Modified starting values q_0, $\mathbf{v}_0$, $\dot{\mathbf{v}}_0$, $\mathbf{a}_0$, $\boldsymbol{\lambda}_0$ of integrator (56)
Step 1	Set starting values q_0, $\mathbf{v}_0$ to the consistent initial values: $q_0 := q(t_0)$, $\mathbf{v}_0 := \mathbf{v}(t_0)$
Step 2	Solve system (19) with $t = t_0$, $q = q_0$, $\mathbf{v} = \mathbf{v}_0$ to get consistent starting values $\dot{\mathbf{v}}_0$ and $\boldsymbol{\lambda}_0$
Step 3	Get $\dot{\mathbf{v}}_{sh}$ from system (19) with $t = t_0 + sh$ and $q = q_0 \circ \exp\big(sh\mathbf{v}_0 + s^2h^2\dot{\mathbf{v}}_0/2\big)$, $\mathbf{v} = \mathbf{v}_0 + sh\dot{\mathbf{v}}_0$
Step 4	Get $\dot{\mathbf{v}}_{-sh}$ from system (19) with $t = t_0 - sh$ and $q = q_0 \circ \exp\big(-sh\mathbf{v}_0 + s^2h^2\dot{\mathbf{v}}_0/2\big)$, $\mathbf{v} = \mathbf{v}_0 - sh\dot{\mathbf{v}}_0$
Step 5	Compute starting value $\mathbf{a}_0 := \dot{\mathbf{v}}_0 + \Delta_\alpha h\,(\dot{\mathbf{v}}_{sh} - \dot{\mathbf{v}}_{-sh})/(2sh)$

of motion at $t_0 + sh$ and at $t_0 - sh$. Then, vectors $\dot{\mathbf{v}}_{sh}$ and $\dot{\mathbf{v}}_{-sh}$ may be obtained from block-structured systems of linear equations (19), see the numerical algorithm in Table 1 for a more detailed discussion of this initialization phase.

Starting values $\mathbf{a}_0$ according to (141) satisfy assumption (137) with $\delta = 1$ since $\dot{\mathbf{v}}(t_0) + \Delta_\alpha h(\dot{\mathbf{v}}_{sh} - \dot{\mathbf{v}}_{-sh})/2sh = \dot{\mathbf{v}}(t_0 + \Delta_\alpha h) + \mathcal{O}(h^2)$. Hence, Theorem 4.18(b) proves second-order convergence of the stabilized index-2 integrator (56) for all solution components. This convergence result may be verified by a numerical test for the heavy top benchmark problem in configuration space $G = \mathrm{SO}(3) \times \mathbb{R}^3$: Fig. 16 shows for time step size $h = 1.0 \times 10^{-3}$ the global error $e_n^{\lambda_1}/\|\boldsymbol{\lambda}_n\|$ of the stabilized index-2 integrator (56) in time interval $[0, 0.1]$. The test results in the left plot are already known from the left plot of Fig. 13. They show the transient oscillating first-order error term being characteristic of the trivial initialization $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0)$. The test results in the right plot illustrate that this first-order error term disappears if we use the modified starting value $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{a}} \approx \dot{\mathbf{v}}(t_0 + \Delta_\alpha h)$.

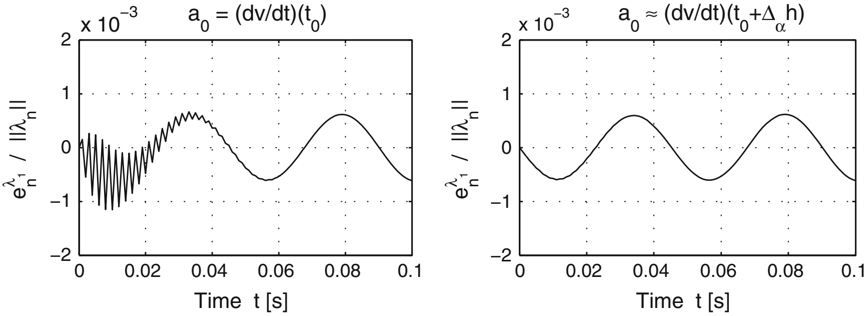


Fig. 16 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, starting values $q_0 = q(t_0)$, $\mathbf{v}_0 = \mathbf{v}(t_0)$, stabilized index-2 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$): Global error $e_n^{\lambda_1}/\|\boldsymbol{\lambda}_n\|$. *Left plot* $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0)$, *right plot* $\mathbf{a}_0 \approx \dot{\mathbf{v}}(t_0 + \Delta_\alpha h)$

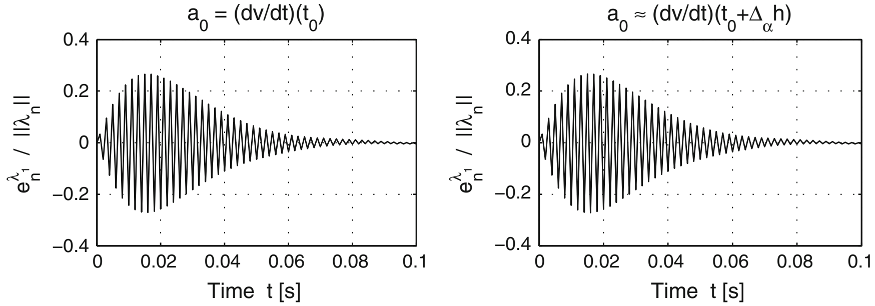


Fig. 17 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, starting values $q_0 = q(t_0)$, $\mathbf{v}_0 = \mathbf{v}(t_0)$, index-3 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$): Global error $e_n^{\lambda_1}/\|\boldsymbol{\lambda}_n\|$. *Left plot* $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0)$, *right plot* $\mathbf{a}_0 \approx \dot{\mathbf{v}}(t_0 + \Delta_\alpha h)$

Note, that this modification of starting value $\mathbf{a}_0$ does not contribute significantly to the result accuracy of the index-3 integrator (37) since the additional assumption (139) in part (a) of Theorem 4.18 is (as before) only satisfied with $\delta = 0$. The resulting large first-order error term in $\lambda_{n,1}$ is (up to plot accuracy) not affected by modified starting values $\mathbf{a}_0$, see Fig. 17.

This first-order error term is well known from the convergence analysis for the linear test problem in Sect. 3.2. In Theorem 3.1(b), we proposed a systematic perturbation of starting values v_0 to get second-order convergence, see (52). In the Lie group setting, these modified starting values are given by

$$\mathbf{v}_0 = \mathbf{v}(t_0) + [\mathbf{M}^{-1}\mathbf{B}^\top(\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^\top)^{-1}\mathbf{B}]\big(q(t_0)\big)\,\mathbf{l}_0^q/h + \mathcal{O}(h^3)\,.$$

In a practical implementation, we restrict ourselves to the leading error term in $\mathbf{l}_0^q$, see (69), and use again a difference approximation of $\ddot{\mathbf{v}}(t_0)$, see (141). The modified starting values are given by $\mathbf{v}_0 = \mathbf{v}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{v}}$ with

$$\begin{aligned}\boldsymbol{\Delta}_0^{\mathbf{v}} := h^2\,[\mathbf{M}^{-1}\mathbf{B}^\top(\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^\top)^{-1}\mathbf{B}]\big(q(t_0)\big)\cdot&\\ \cdot\Big(C_q\,\frac{\dot{\mathbf{v}}_{sh} - \dot{\mathbf{v}}_{-sh}}{2sh} + \frac{1}{12}\widehat{\mathbf{v}}(t_0)\dot{\mathbf{v}}(t_0)\Big)\,.&\end{aligned} \tag{142}$$

They may be computed efficiently by the numerical algorithm in Table 2. The numerical test results for two different time step sizes in Fig. 18 illustrate that the modified starting values eliminate the first-order error term. The maximum amplitude of $e_n^{\lambda_1}/\|\boldsymbol{\lambda}_n\|$ is reduced by a factor of 4 if the time step size is reduced from $h = 1.0 \times 10^{-3}$ to $h = 5.0 \times 10^{-4}$.

The perturbation of size $\mathcal{O}(h^2)$ in (142) results in starting values $q_0 = q(t_0)$ and $\mathbf{v}_0 = \mathbf{v}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{v}}$ that satisfy assumption (139) in Theorem 4.18(a) with $\delta = 1$. In general, these starting values are *not* consistent with the hidden constraints (16) at the level of velocity coordinates but introduce systematically a residual of size $\mathbf{B}(q_0)\mathbf{v}_0 = \mathcal{O}(h^2)$ at $t = t_0$. The numerical test results in Figs. 19 and 20 show that this

Table 2 Initialization of the index-3 integrator (37)

Data	Consistent initial values $q(t_0)$, $\mathbf{v}(t_0)$; parameter $s \in (0, 1]$
Result	Modified starting values q_0, $\mathbf{v}_0$, $\dot{\mathbf{v}}_0$, $\mathbf{a}_0$, $\boldsymbol{\lambda}_0$ of integrator (37)
Step 1	Set starting value q_0 to the consistent initial value: $q_0 := q(t_0)$
Step 2	Solve system (19) with $t = t_0$, $q = q(t_0)$, $\mathbf{v} = \mathbf{v}(t_0)$ to get consistent starting values $\dot{\mathbf{v}}_0$ and $\boldsymbol{\lambda}_0$
Step 3	Get $\dot{\mathbf{v}}_{sh}$ from system (19) with $t = t_0 + sh$ and $q = q(t_0) \circ \exp\big(sh\mathbf{v}(t_0) + s^2h^2\dot{\mathbf{v}}_0/2\big)$, $\mathbf{v} = \mathbf{v}(t_0) + sh\dot{\mathbf{v}}_0$
Step 4	Get $\dot{\mathbf{v}}_{-sh}$ from system (19) with $t = t_0 - sh$ and $q = q(t_0) \circ \exp\big(-sh\mathbf{v}(t_0) + s^2h^2\dot{\mathbf{v}}_0/2\big)$, $\mathbf{v} = \mathbf{v}(t_0) - sh\dot{\mathbf{v}}_0$
Step 5	Compute starting value $\mathbf{a}_0 := \dot{\mathbf{v}}_0 + \Delta_\alpha h\,(\dot{\mathbf{v}}_{sh} - \dot{\mathbf{v}}_{-sh})/(2sh)$
Step 6	Get $\boldsymbol{\Delta}_0^{\mathbf{v}} := \mathbf{x}_{\dot{\mathbf{v}}}$ from the system of linear equations (20) with $\mathbf{r}_{\dot{\mathbf{v}}} = \mathbf{0}_k\,,\quad \mathbf{r}_{\boldsymbol{\lambda}} = h^2\mathbf{B}(q_0)\big(C_q\dfrac{\dot{\mathbf{v}}_{sh} - \dot{\mathbf{v}}_{-sh}}{2sh} + \dfrac{1}{12}\widehat{\mathbf{v}}(t_0)\dot{\mathbf{v}}(t_0)\big)$ and matrices $\mathbf{M} = \mathbf{M}(q_0)$, $\mathbf{B} = \mathbf{B}(q_0)$
Step 7	Set starting value $\mathbf{v}_0$ to $\mathbf{v}_0 := \mathbf{v}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{v}}$, see (142)

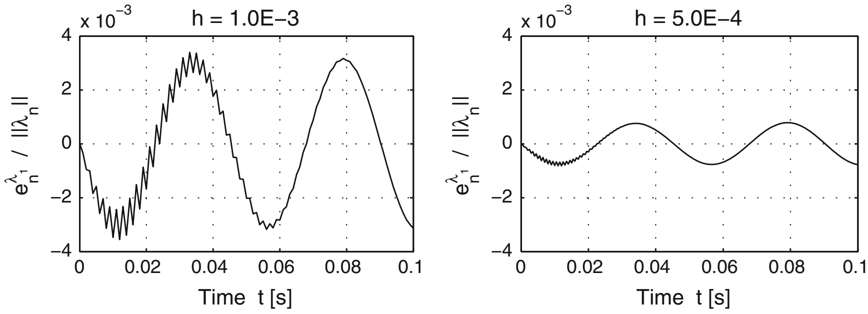


Fig. 18 Heavy top benchmark (index-3 formulation, starting values $q_0 = q(t_0)$, $\mathbf{v}_0 = \mathbf{v}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{v}}$, $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{a}}$, $G = \mathrm{SO}(3) \times \mathbb{R}^3$): Global error $e_n^{\lambda_1}/\|\boldsymbol{\lambda}_n\|$. *Left plot* $h = 1.0 \times 10^{-3}$, *right plot* $h = 5.0 \times 10^{-4}$

non-vanishing initial constraint residual helps to avoid the oscillating second-order term in the constraint residuals $\mathbf{B}(q_n)\mathbf{v}_n$ as well as the corresponding oscillating first-order error term in the Lagrange multipliers $\boldsymbol{\lambda}_n$: In the left plots of Figs. 19 and 20, we see the simulation data for (classical) starting values $\mathbf{v}_0 = \mathbf{v}(t_0)$, $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0)$ that are already known from the numerical tests in Sect. 3.3 (left plots of Figs. 10 and 7). The test results in the right plots of Figs. 19 and 20 show that the transient oscillating terms disappear up to plot accuracy for the modified starting values $\mathbf{v}_0 = \mathbf{v}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{v}} = \mathbf{v}(t_0) + \mathcal{O}(h^2)$ and $\mathbf{a}_0 = \dot{\mathbf{v}}(t_0) + \boldsymbol{\Delta}_0^{\mathbf{a}} = \dot{\mathbf{v}}(t_0 + \Delta_\alpha h) + \mathcal{O}(h^2)$.

The algorithm in Table 2 spends moderate numerical effort to get (modified) starting values q_0, $\mathbf{v}_0$, $\dot{\mathbf{v}}_0$, $\mathbf{a}_0$ and $\boldsymbol{\lambda}_0$ for the generalized-α Lie group integrator (37) that satisfy assumptions (137) and (139) in the convergence theorem with $\delta = 1$. The error bounds (138) in Theorem 4.18(a) prove second-order convergence in all solu-

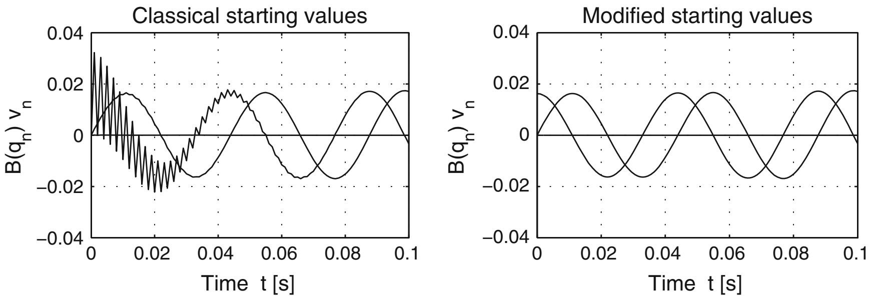


Fig. 19 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, index-3 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$): Residuals in hidden constraints (16). *Left plot* classical starting values $\mathbf{v}_0$, $\mathbf{a}_0$, *right plot* modified starting values $\mathbf{v}_0$, $\mathbf{a}_0$

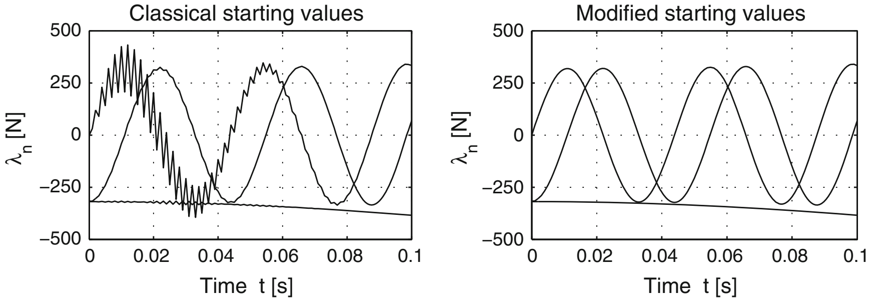


Fig. 20 Heavy top benchmark ($h = 1.0 \times 10^{-3}$, index-3 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$): Numerical solution $\boldsymbol{\lambda}_n$. *Left plot* classical starting values $\mathbf{v}_0$, $\mathbf{a}_0$, *right plot* modified starting values $\mathbf{v}_0$, $\mathbf{a}_0$

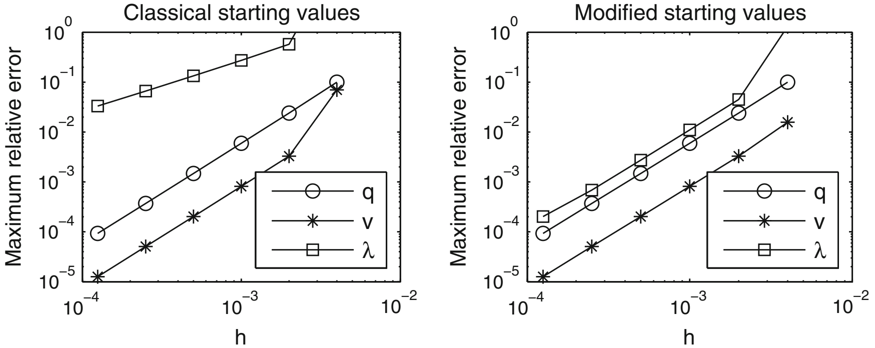


Fig. 21 Heavy top benchmark (index-3 formulation, $G = \mathrm{SO}(3) \times \mathbb{R}^3$): Global error of integrator (37) versus h for $t \in [0, 1]$. *Left plot* classical starting values $\mathbf{v}_0$, $\mathbf{a}_0$, *right plot* modified starting values $\mathbf{v}_0$, $\mathbf{a}_0$

tion components. The right plot of Fig. 21 shows numerical test results for the heavy top benchmark problem that are in perfect agreement with this asymptotic error analysis for small time step sizes h.

5 Summary

The generalized-α method is a Newmark-type method and one of the standard time integration methods in structural dynamics. The method is second-order accurate for unconstrained systems in linear spaces and has a free algorithmic parameter that allows to control the amount of numerical dissipation for high-frequency solution components. Following a Lie algebra approach, the method may be applied as well to mechanical systems that have a nonlinear configuration space with Lie group structure. In each time step, the increment of the configuration variables is parametrized by an element of the corresponding Lie algebra that may be obtained numerically by a classical Newton–Raphson iteration in linear spaces.

The Lie algebra approach is used as well in the asymptotic error analysis for the application to constrained systems that are typical of multibody dynamics. Newmark-type time integration methods of second-order accuracy are known to suffer from "overshooting", i.e., from an oscillating transient error term in the application to a scalar linear test equation with high-frequency solutions. For constrained systems, these large transient errors may result in order reduction unless the starting values of the generalized-α method are perturbed by an appropriate second-order correction term. Second-order convergence of the algorithm with perturbed starting values is proved analytically studying a coupled error propagation in differential and algebraic solution components that takes into account a quadratic approximation of hidden constraints at the level of acceleration coordinates.

The order reduction phenomenon may be avoided by an analytical index reduction before time discretization. The Lie algebra approach allows to modify the increment of configuration variables such that the numerical solution satisfies in each time step the original holonomic constraints at the level of position coordinates as well as the corresponding hidden constraints at the level of velocity coordinates (stabilized index-2 formulation). With an appropriate initialization of the acceleration like variables $\mathbf{a}_n$ in the generalized-α method, this stabilized index-2 Lie group DAE integrator is second-order accurate for any starting values being consistent with original and hidden constraints in the equations of motion.

All results of the convergence analysis have been verified in detail by numerical tests for a heavy top benchmark problem in Lie groups $\mathrm{SO}(3) \times \mathbb{R}^3$ and $\mathrm{SE}(3)$, respectively. The theoretical investigations are limited to fixed time step sizes but will be extended to variable step size implementations with error control in future work. In that case, the acceleration like variables $\mathbf{a}_n$ need to be updated whenever the time step size is changed at $t = t_n$. Furthermore, the velocity vector $\mathbf{v}_n$ has to be perturbed by an appropriate second-order correction term unless the generalized-α Lie group DAE integrator is applied to the index-reduced stabilized index-2 formulation of the equations of motion.

Acknowledgments The work of this author was supported by the German Minister of Education and Research, BMBF grant 05M13NHB: "Modelling and structure preserving discretization of inelastic components in system simulation".

References

Arnold, M., & Brüls, O. (2007). Convergence of the generalized-α scheme for constrained mechanical systems. *Multibody System Dynamics*, *18*, 185–202. doi:10.1007/s11044-007-9084-0.

Arnold, M., Brüls, O., & Cardona, A. (2011a). Convergence analysis of generalized-α Lie group integrators for constrained systems. In J. C. Samin & P. Fisette (Eds.), *Proceedings of Multibody Dynamics 2011 (ECCOMAS Thematic Conference)*, Brussels, Belgium.

Arnold, M., Brüls, O., & Cardona, A. (2011b). Improved stability and transient behaviour of generalized-α time integrators for constrained flexible systems. In *Fifth International Conference on Advanced COmputational Methods in ENgineering (ACOMEN 2011)*, Liège, Belgium, 14–17 November 2011.

Arnold, M., Burgermeister, B., Führer, C., Hippmann, G., & Rill, G. (2011c). Numerical methods in vehicle system dynamics: State of the art and current developments. *Vehicle System Dynamics*, *49*, 1159–1207. doi:10.1080/00423114.2011.582953.

Arnold, M., Cardona, A., & Brüls, O. (2014). Order reduction in time integration caused by velocity projection. In *Proceedings of the 3rd Joint International Conference on Multibody System Dynamics and the 7th Asian Conference on Multibody Dynamics*, 30 June–3 July 2014. *BEXCO*. Korea: Busan.

Arnold, M., Brüls, O., & Cardona, A. (2015). Error analysis of generalized-α Lie group time integration methods for constrained mechanical systems. *Numerische Mathematik*, *129*, 149–179. doi:10.1007/s00211-014-0633-1.

Betsch, P., & Leyendecker, S. (2006). The discrete null space method for the energy consistent integration of constrained mechanical systems. Part II: Multibody dynamics. *International Journal for Numerical Methods in Engineering*, *67*, 499–552. doi:10.1002/nme.1639.

Betsch, P., & Siebert, R. (2009). Rigid body dynamics in terms of quaternions: Hamiltonian formulation and conserving numerical integration. *International Journal for Numerical Methods in Engineering*, *79*, 444–473. doi:10.1002/nme.2586.

Bottasso, C. L., & Borri, M. (1998). Integrating finite rotations. *Computer Methods in Applied Mechanics and Engineering*, *164*, 307–331. doi:10.1016/S0045-7825(98)00031-0.

Bottasso, C. L., Bauchau, O. A., & Cardona, A. (2007). Time-step-size-independent conditioning and sensitivity to perturbations in the numerical solution of index three differential algebraic equations. *SIAM Journal on Scientific Computing*, *29*, 397–414. doi:10.1137/050638503.

Brenan, K. E., Campbell, S. L., & Petzold, L. R. (1996). *Numerical solution of initial-value problems in differential-algebraic equations* (2nd ed.). Philadelphia: SIAM.

Brüls, O., & Arnold, M. (2008). The generalized-α scheme as a linear multistep integrator: Towards a general mechatronic simulator. *Journal of Computational and Nonlinear Dynamics*, *3*(4), 041007. doi:10.1115/1.2960475.

Brüls, O., & Cardona, A. (2010). On the use of Lie group time integrators in multibody dynamics. *Journal of Computational and Nonlinear Dynamics*, *5*, 031002. doi:10.1115/1.4001370.

Brüls, O., & Golinval, J. C. (2006). The generalized-α method in mechatronic applications. *ZAMM—Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, *86*, 748–758. doi:10.1002/zamm.200610283.

Brüls, O., Arnold, M., & Cardona, A. (2011). Two Lie group formulations for dynamic multibody systems with large rotations. In *Proceedings of IDETC/MSNDC 2011, ASME 2011 International Design Engineering Technical Conferences*, Washington, USA.

Brüls, O., Cardona, A., & Arnold, M. (2012). Lie group generalized-α time integration of constrained flexible multibody systems. *Mechanism and Machine Theory*, *48*, 121–137. doi:10.1016/j.mechmachtheory.2011.07.017.

Cardona, A., & Géradin, M. (1989). Time integration of the equations of motion in mechanism analysis. *Computers and Structures*, *33*, 801–820. doi:10.1016/0045-7949(89)90255-1.

Cardona, A., & Géradin, M. (1994). Numerical integration of second order differential-algebraic systems in flexible mechanism dynamics. In M. F. O. Seabra Pereira & J. A. C. Ambrósio (Eds.), *Computer-Aided Analysis of Rigid and Flexible Mechanical Systems*, *NATO ASI Series* (Vol. E–268). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-011-1166-9_16.

Celledoni, E., & Owren, B. (2003). Lie group methods for rigid body dynamics and time integration on manifolds. *Computer Methods in Applied Mechanics and Engineering*, *192*, 421–438. doi:10.1016/S0045-7825(02)00520-0.

Chung, J., & Hulbert, G. (1993). A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized-α method. *ASME Journal of Applied Mechanics*, *60*, 371–375. doi:10.1115/1.2900803.

Crouch, P. E., & Grossman, R. (1993). Numerical integration of ordinary differential equations on manifolds. *Journal of Nonlinear Science*, *3*, 1–33. doi:10.1007/BF02429858.

Deuflhard, P., Hairer, E., & Zugck, J. (1987). One-step and extrapolation methods for differential-algebraic systems. *Numerische Mathematik*, *51*, 501–516. doi:10.1007/BF01400352.

Erlicher, S., Bonaventura, L., & Bursi, O. (2002). The analysis of the generalized-α method for non-linear dynamic problems. *Computational Mechanics*, *28*, 83–104. doi:10.1007/s00466-001-0273-z.

García de Jalón, J., & Bayo, E. (1994). *Kinematic and dynamic simulation of multibody systems: The real-time challenge*. New York: Springer-Verlag.

García de Jalón, J., & Gutiérrez-López, M. D. (2013). Multibody dynamics with redundant constraints and singular mass matrix: Existence, uniqueness, and determination of solutions for accelerations and constraint forces. *Multibody System Dynamics*, *30*, 311–341. doi:10.1007/s11044-013-9358-7.

Gear, C. W., Leimkuhler, B., & Gupta, G. K. (1985). Automatic integration of Euler-Lagrange equations with constraints. *Journal of Computational and Applied Mathematics*, *12&13*, 77–90. doi:10.1016/0377-0427(85)90008-1.

Géradin, M., & Cardona, A. (2001). *Flexible multibody dynamics: A finite element approach*. Chichester: Wiley.

Géradin, M., & Cardona, A. (1989). Kinematics and dynamics of rigid and flexible mechanisms using finite elements and quaternion algebra. *Computational Mechanics*, *4*, 115–135. doi:10.1007/BF00282414.

Golub, G. H., & van Loan, Ch. F. (1996). *Matrix computations* (3rd ed.). Baltimore London: The Johns Hopkins University Press.

Hairer, E., & Wanner, G. (1996). *Solving ordinary differential equations. II. Stiff and differential-algebraic problems* (2nd ed.). Berlin, Heidelberg, New York: Springer-Verlag.

Hairer, E., Nørsett, S. P., & Wanner, G. (1993). *Solving ordinary differential equations. I. Nonstiff problems* (2nd ed.). Berlin, Heidelberg, New York: Springer-Verlag.

Hairer, E., Lubich, Ch., & Wanner, G. (2006). *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations* (2nd ed.). Berlin, Heidelberg, New York: Springer-Verlag.

Hilber, H. M., & Hughes, T. J. R. (1978). Collocation, dissipation and 'overshoot' for time integration schemes in structural dynamics. *Earthquake Engineering and Structural Dynamics*, *6*, 99–117. doi:10.1002/eqe.4290060111.

Hilber, H. M., Hughes, T. J. R., & Taylor, R. L. (1977). Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthquake Engineering and Structural Dynamics*, *5*, 283–292. doi:10.1002/eqe.4290050306.

Iserles, A., Munthe-Kaas, H. Z., Nørsett, S., & Zanna, A. (2000). Lie-group methods. *Acta Numerica*, *9*, 215–365.

Jay, L. O., & Negrut, D. (2007). Extensions of the HHT-method to differential-algebraic equations in mechanics. *Electronic Transactions on Numerical Analysis*, *26*, 190–208.

Jay, L. O., & Negrut, D. (2008). A second order extension of the generalized-α method for constrained systems in mechanics. In C. Bottasso (Ed.), *Multibody Dynamics. Computational Methods and Applications*, *Computational Methods in Applied Sciences* (Vol. 12, pp. 143–158). Dordrecht: Springer. doi:10.1007/978-1-4020-8829-2_8.

Kelley, C. T. (1995). *Iterative methods for linear and nonlinear equations*. Philadelphia: SIAM.

Kobilarov, M., Crane, K., & Desbrun, M. (2009). Lie group integrators for animation and control of vehicles. *ACM Transactions on Graphics*, *28*(2, Article 16), 1–14. doi:10.1145/1516522.1516527.

Lubich, Ch. (1993). Integration of stiff mechanical systems by Runge-Kutta methods. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, *44*, 1022–1053. doi:10.1007/BF00942763.

Lunk, C., & Simeon, B. (2006). Solving constrained mechanical systems by the family of Newmark and α-methods. *Zeitschrift für Angewandte Mathematik und Mechanik*, *86*, 772–784. doi:10.1002/zamm.200610285.

Müller, A. (2010). Approximation of finite rigid body motions from velocity fields. *ZAMM—Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, *90*, 514–521. doi:10.1002/zamm.200900383.

Müller, A., & Terze, Z. (2014a). On the choice of configuration space for numerical Lie group integration of constrained rigid body systems. *Journal of Computational and Applied Mathematics*, *262*, 3–13. doi:10.1016/j.cam.2013.10.039.

Müller, A., & Terze, Z. (2014b). The significance of the configuration space Lie group for the constraint satisfaction in numerical time integration of multibody systems. *Mechanism and Machine Theory*, *82*, 173–202. doi:10.1016/j.mechmachtheory.2014.06.014.

Munthe-Kaas, H. (1995). Lie-Butcher theory for Runge-Kutta methods. *BIT Numerical Mathematics*, *35*, 572–587. doi:10.1007/BF01739828.

Munthe-Kaas, H. (1998). Runge-Kutta methods on Lie groups. *BIT Numerical Mathematics*, *38*, 92–111. doi:10.1007/BF02510919.

Negrut, D., Rampalli, R., Ottarsson, G., & Sajdak, A. (2005). On the use of the HHT method in the context of index 3 differential algebraic equations of multi-body dynamics. In J. M. Goicolea, J. Cuadrado & J. C. García Orden (Eds.), *Proceedings of Multibody Dynamics 2005 (ECCOMAS Thematic Conference)*, Madrid, Spain.

Orel, B. (2010). Accumulation of global error in Lie group methods for linear ordinary differential equations. *Electronic Transactions on Numerical Analysis*, *37*, 252–262.

Petzold, L. R., & Lötstedt, P. (1986). Numerical solution of nonlinear differential equations with algebraic constraints II: Practical implications. *SIAM Journal on Scientific and Statistical Computing*, *7*, 720–733. doi:10.1137/0907049.

Quarteroni, A., Sacco, R., & Saleri, F. (2000). *Numerical mathematics*. New York: Springer.

Sanborn, G. G., Choi, J., & Choi, J. H. (2014). Review of RecurDyn integration methods. In Proceedings of the 3rd Joint International Conference on Multibody System Dynamics and the 7th Asian Conference on Multibody Dynamics, 30 June–3 July (2014). *BEXCO*. Korea: Busan.

Simo, J. C., & Vu-Quoc, L. (1988). On the dynamics in space of rods undergoing large motions—A geometrically exact approach. *Computer Methods in Applied Mechanics and Engineering*, *66*, 125–161. doi:10.1016/0045-7825(88)90073-4.

Sonneville, V., Cardona, A., & Brüls, O. (2014). Geometrically exact beam finite element formulated on the special Euclidean group. *Computer Methods in Applied Mechanics and Engineering*, *268*, 451–474. doi:10.1016/j.cma.2013.10.008.

Walter, W. (1998). *Ordinary Differential Equations*. Number 182 in Graduate Texts in Mathematics. Berlin: Springer.

Wensch, J. (2001). Extrapolation methods in Lie groups. *Numerische Mathematik*, *89*, 591–604. doi:10.1007/s211-001-8017-5.

The Absolute Nodal Coordinate Formulation

Johannes Gerstmayr, Alexander Humer, Peter Gruber and Karin Nachbagauer

Abstract The key idea of the absolute nodal coordinate formulation (ANCF) is to use slope vectors in order to describe the orientation of the cross-section of structural mechanics components, such as beams, plates or shells. This formulation relaxes the kinematical assumptions of Bernoulli–Euler and Timoshenko beam theories and enables a deformation of the cross-sections. The present contribution shows how to create 2D and 3D structural finite elements based on the ANCF by employing different sets of slope vectors for approximating the cross-sections' orientation. A specific aim of this chapter is to present a unified notation for structural mechanics and continuum mechanics ANC formulations. Particular focus is laid on enhanced formulations for such finite elements that circumvent severe issues like Poisson or shear locking. The performance of these elements is evaluated and a detailed assessment comprising the convergence order, the number of iterations, and Jacobian updates for large deformation benchmark problems is provided.

1 Introduction

In a world with an increasing amount of automation, mobility, adaptive structures, and miniaturized systems, the modeling and simulation of flexible multibody systems gains importance. Large deformation of some components can significantly influence

J. Gerstmayr (✉)
Leopold-Franzens-Universität Innsbruck, Technikerstraße 13,
6020 Innsbruck, Austria
e-mail: johannes.gerstmayr@uibk.ac.at

A. Humer · P. Gruber
Linz Center of Mechatronics GmbH, Altenberger Straße 69, 4040 Linz, Austria

A. Humer
Johannes Kepler University, Altenberger Straße 69, 4040 Linz, Austria

K. Nachbagauer
University of Applied Sciences Upper Austria, Stelzhamerstraße 23,
4600 Wels, Austria

P. Betsch (ed.), *Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics*, CISM International Centre for Mechanical Sciences 565, DOI 10.1007/978-3-319-31879-0_4

the behavior of the flexible multibody system. Examples are the dynamics of thin rotor blades, transportation of sheets or strips, various kinds of cables, wires, and tires.

There are several possibilities to study the dynamic behavior of slender structures. A convenient way to model large deformations of beam-like structures is to combine several beams described by the floating frame of reference formulation with an individual frame for each beam. As soon as the number of beams becomes larger, the solution of geometrically nonlinear problems converges to the solution of nonlinear beam formulations, see Gerstmayr and Irschik (2003) and Dibold et al. (2009). The floating frame approach has drawbacks like inappropriate modeling of nonlinearities for geometric stiffening and slow convergence and it cannot be extended to shells. Furthermore, the equations of motion as well as the constraint conditions for pairwise interconnection of beams become tedious. In finite element codes, large deformation structural finite elements based on the large rotation vector formulation of Simo and Vu-Quoc (1988) are available for studying the dynamics of thin structures. These elements require special time integration methods for stable long-term dynamic simulations.

In the present chapter, we focus on beam finite elements based on the absolute nodal coordinate formulation. Specifically, the focus of this chapter lies on a class of thin beam finite elements, based on the Bernoulli–Euler beam theory, and a class of thick beam finite elements, which include shear and cross-section deformation. This chapter provides a brief overview of existing absolute nodal coordinate (ANC) formulations, relations to other modeling techniques for large deformation beam finite elements, details on the formulation and implementation of the equations of motion, and some representative numerical tests that show the order of convergence, the performance and the stability of ANC beam finite elements.

1.1 ANCF—Basic Ideas

This section aims to highlight various basic ideas for ANC finite elements. For a recent review article on ANCF, which provides important references, see Gerstmayr et al. (2013b). We like to emphasize that some of the subsequent ideas do not apply to every ANC finite element published in the literature. In addition to that, there is no general definition whether to call a finite element ANC element, or not.

The first, and probably most widely accepted, idea is that ANC finite elements are based on slope vectors[1] rather than rotation parameters such as Euler angles or Euler parameters. Rotational parameters can immediately lead to a numerically induced blow up of the total energy in a conservative flexible multibody system, see the examples section of this chapter as well as the classical literature on 3D nonlinear beam formulations of the 1980s and 1990s, see Simo and Vu-Quoc (1988). As an advantage of the ANCF, slope vectors can be interpolated in space and time in

[1] For an example of a slope vector, see $\mathbf{x}_{,\xi}$, $\mathbf{x}_{,\eta}$ or $\mathbf{x}_{,\zeta}$ in Fig. 2.

the same way as displacements, which does not lead to well-known problems of interpolation of rotations. As a disadvantage, the slope vectors are stiffly constrained to the nearly-rigid-body motion of the cross-section, which can cause high-frequency dynamics behavior.

As a result of a pure displacement (or displacement gradient) interpolation in space, ANC finite elements usually employ a constant mass matrix. This can lead to simpler implementation and computational efficiency. The straightforward kinematic description of the motion of each point of the beam makes an extension to advanced kinematics descriptions (such as ALE) or to multi-physics coupling much easier, see Pechstein and Gerstmayr (2013).

From the computational point of view, ANC finite elements are solved according to a Total Lagrangian (TL) scheme. This means, that no incremental (or co-rotational) formulation is utilized, which is sometimes applied in formulations based on rotational parameters.

ANC finite elements shall be capable of large deformations (in comparison to structural finite elements based on the floating frame of reference formulation) and can even be applied to (moderately) large strains. Specifically, in some sort of ANC finite elements, 3D continuum mechanics material laws can be directly applied, which makes this formulation attractive, e.g., for rubber-like materials, see Irschik and Gerstmayr (2009a).

The original shear and cross-section deformable ANC finite elements relax the assumptions of the classical Bernoulli–Euler and the Timoshenko beam theory, in the sense that the cross-section is not rigid any longer. As a consequence, the shrinkage of parts of the cross-section due to elongation can be modeled, which has many applications, e.g., in rolling processes.

In the case of so-called fully parametrized ANC finite elements, which use three slope vectors for the definition of the orientation of the cross-section, an interconnection of finite elements at any angle is possible without the need of constraint conditions, see Sugiyama et al. (2003).

There is a general transformation of the NURBS-based geometry of slender structures to ANC finite elements, which allows the direct computation of CAD geometry without the need for an intermediate discretization, see Lan and Shabana (2010a, b).

1.2 ANCF—Short Summary

There exist a vast amount of structural finite elements in the literature. Many of the proposed structural finite elements have specific objectives and purposes. Among other things, ANC finite elements have been designed for simulation of the dynamics of flexible multibody systems consisting of structural components. In this context, the term “structural” is used in order to distinguish such elements from conventional solid finite elements.

In one of the earliest papers on ANCF, Escalona et al. (1998) proposed a polynomial interpolation of the position of the beam axis for the computation of the deformation energy, the kinetic energy and the mass matrix. In the latter paper, the authors used a planar Bernoulli–Euler beam theory, using a cubic interpolation along the axis of the beam finite element. A co-rotational frame is defined, which is spanned by the end points of the beam finite element, in order to compute the strain energy. However, the mass matrix becomes constant and the formulation can be implemented very efficiently. In order to extend the latter idea, it is possible to use cross-section slope vectors, see Yakoub and Shabana (2001), and to use a similar co-rotational linearization, see, e.g., Gerstmayr (2009).

The absolute nodal coordinate formulation facilitates the application of constitutive relations on the continuum mechanics level, therefore, almost arbitrary material laws as well as large strain formulations can be incorporated in a straightforward manner. The classical large deformation beam finite elements, which have been proposed by Simo (1985) and Simo and Vu-Quoc (1986c), are based on a strain energy which is a quadratic function of generalized strain measures such as axial strain or curvature. These strain measures can be interpreted in terms of continuum mechanics quantities, see Irschik and Gerstmayr (2009b), however, the ANCF allows for a much simpler realization of nonlinear, e.g., hyperelastic, material laws, see Irschik and Gerstmayr (2009a).

There are other approaches than ANC finite elements for the combination of continuum mechanics with structural finite elements, see Frischkorn and Reese (2012) for a recent work on beams modeled with hexahedrals. The slope vectors in the ANCF can be directly related to well-known director based methods, if constraints are applied to the length of the slopes vectors and for the orthogonality of the slope vectors. Applying constraints on a fully parametrized ANC beam finite element is in line with the approach proposed by Betsch and Steinmann (2003), see the corresponding chapter in this book. Furthermore, the latter approach is based on the geometrically exact beam formulation of Simo (1985).

There is one important group of so-called fully parametrized ANC finite elements. The term 'fully parameterized' indicates that all nine components of the spatial deformation gradient (four in the planar case) are used as coordinates in each node. Using these coordinates, it is possible to interconnect ANC finite elements with slope discontinuities without any constraint equations, see Sugiyama et al. (2003). In this chapter, a specific focus is laid on so-called gradient-deficient ANC finite elements, which means that less slope vectors are used than in the fully parametrized case.

There are several issues concerning the ANCF, which are not discussed in detail in the present chapter. The idea of using slopes as nodal degrees of freedom has been extended to plates, see Mikkola and Shabana (2003), resp. shells and general 3D solids, see Olshevskiy et al. (2013). In the present chapter, we only discuss planar and spatial ANC beam finite elements. The continuum mechanics formulation, which is frequently used for the computation of the elastic forces in ANCF, is well suited for the modeling of nonlinear elastic material, see Irschik and Gerstmayr (2011), or inelastic material behavior, see Sugiyama and Shabana (2004) and Gerstmayr and Matikainen (2006). The latter topics are not addressed in the present chapter.

2 General Formulation of ANC Beam Elements

2.1 Kinematics of ANC Beam Elements

In the present section, the kinematic preliminaries describing the deformation of beams are introduced. Throughout the following sections, we employ a direct tensor notation or tensor components where appropriate. Einstein's notion of summation over repeated indices is used for the sake of brevity. The scalar product of two vectors is given as $\mathbf{a}^T\mathbf{b} = a_i b_i$. The composition of two tensors and the linear mapping of a vector by a tensor read $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ and $\mathbf{A}\mathbf{b} = A_{ij}b_j$, respectively. The double contraction of tensors is indicated by a colon, e.g., the inner product of second-order tensors, i.e., $\mathbf{A} : \mathbf{B} = \mathrm{tr}(\mathbf{A}^T\mathbf{B}) = A_{ij}B_{ij}$; the product of a fourth- and a second-order tensor is defined as ${}^4\mathbf{C} : \mathbf{A} = C_{ijkl}A_{kl}$. For the tensor product of two vectors, we use the notation $\mathbf{a} \otimes \mathbf{b} = a_i b_j$.

Regardless of whether structural finite elements as beams and plates or conventional solid elements are considered, large deformation problems in continuum mechanics require an exact representation of the geometry of deformation. As is customary in solid mechanics, a reference configuration is introduced which primarily serves the purpose of identifying a body's material points. In the material or Lagrangian representation employed subsequently, the field variables are functions of the material points, or rather, their positions in the reference configuration. In order to avoid curvilinear coordinates, the reference configuration—not necessarily occupied by the body in the course of deformation—is a straight beam whose axis is aligned with the x-axis of some fixed Cartesian frame $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$. Let (ξ, η, ζ) denote the (straight) referential coordinates, see Fig. 1, such that the position of some point P is identified by the vector $\boldsymbol{\xi}$,

$$\boldsymbol{\xi} = \xi\mathbf{e}_x + \eta\mathbf{e}_y + \zeta\mathbf{e}_z. \tag{1}$$

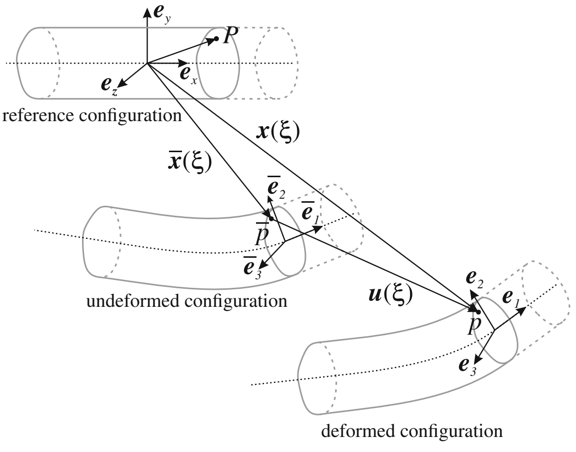


Fig. 1 Important geometrical definitions for ANC elements

In general, the undeformed beam can be curved and arbitrarily oriented relative to the previously introduced fixed frame. The undeformed configuration, relative to which the deformation is measured, therefore has to be distinguished from the reference configuration. The position of the material point P in the undeformed configuration is denoted by $\bar{\mathbf{x}}$, whose coordinates $(\bar{x}, \bar{y}, \bar{z})$ relative to the fixed frame are referred to as material coordinates subsequently:

$$\bar{\mathbf{x}} = \bar{x}\mathbf{e}_x + \bar{y}\mathbf{e}_y + \bar{z}\mathbf{e}_z. \tag{2}$$

The position in the undeformed state is related to the current position in the deformed configuration $\mathbf{x}$ by means of the displacement vector $\mathbf{u}$, i.e.,

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{u} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z. \tag{3}$$

The position vector to a material point of the beam's axis in undeformed configuration is defined as $\bar{\mathbf{r}}$ whereas in deformed configuration it reads $\mathbf{r}$.

Besides the idea of cross-sectional stress resultants, restrictions concerning the deformation of the beam's cross-section are a key ingredient enabling a reduction of a 3D problem to a 1D problem of a beam. All the beam finite elements discussed subsequently can be considered as more or less special cases of a single set of kinematic assumptions: Cross-sections, initially plane and perpendicular to the beam's axis in the undeformed configuration, remain plane in the course of deformation. In contrast to conventional formulations, however, we want to allow the cross-sections to change their size and shape, i.e., a constant in-plane stretch and shearing. Timoshenko's hypothesis would be recovered by prohibiting the latter; the classical assumption for slender structures attributed to Bernoulli and Euler would be obtained by further restricting that the cross-sections remain perpendicular to the beam's axis during deformation.

In the most general case considered herein, the position of the material point P can therefore be expressed in terms of the axis' initial and current position, i.e., $\bar{\mathbf{r}}$ and $\mathbf{r}$, respectively, as

$$\mathbf{x} = \mathbf{r} + \mathbf{A}\,(\bar{\mathbf{x}} - \bar{\mathbf{r}}) + \mathbf{u}_{\mathrm{cs}}, \tag{4}$$

where $\mathbf{u}_{\mathrm{cs}}$ denotes the in-plane deformation of the cross-sections and the second-order tensor $\mathbf{A}$ represents the rotation of the local frame in P from the undeformed to the deformed configuration:

$$\mathbf{A} = \mathbf{e}_i \otimes \bar{\mathbf{e}}_i. \tag{5}$$

The notions of bending and shear deformation in beam theories are intrinsically related to body-local directions. In order to specify the strain measures the subsequent formulations are based on, we therefore need to specify local frames in the beam's configurations used in the analysis. As the beam is straight in the reference configuration, we choose the associated local frame in the directions of the global Cartesian frame $\{\mathbf{e}_{\mathrm{ref},1} = \mathbf{e}_x, \mathbf{e}_{\mathrm{ref},2} = \mathbf{e}_y, \mathbf{e}_{\mathrm{ref},3} = \mathbf{e}_z\}$. Expressing the position in the undeformed configuration in terms of the referential coordinates, $\bar{\mathbf{x}} = \bar{\mathbf{x}}(\xi, \eta, \zeta)$, the corresponding local (cross-section) frame may be defined in dependence of the lateral slope vectors $\bar{\mathbf{x}}_{,\eta}$ and $\bar{\mathbf{x}}_{,\zeta}$. Under the convenient assumption that the undeformed configuration is chosen such that $\bar{\mathbf{x}}_{,\eta}$ and $\bar{\mathbf{x}}_{,\zeta}$ are perpendicular at every $\bar{\mathbf{x}}$, the definition of the local frame $(\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3)$ reads

$$\bar{\mathbf{e}}_1 = \frac{\bar{\mathbf{x}}_{,\eta} \times \bar{\mathbf{x}}_{,\zeta}}{\|\bar{\mathbf{x}}_{,\eta} \times \bar{\mathbf{x}}_{,\zeta}\|}, \quad \bar{\mathbf{e}}_2 = \frac{\bar{\mathbf{x}}_{,\eta}}{\|\bar{\mathbf{x}}_{,\eta}\|}, \quad \bar{\mathbf{e}}_3 = \frac{\bar{\mathbf{x}}_{,\zeta}}{\|\bar{\mathbf{x}}_{,\zeta}\|}. \tag{6}$$

Apparently, $\bar{\mathbf{e}}_1$ is perpendicular to the undeformed cross-section, whereas $\bar{\mathbf{e}}_2$ and $\bar{\mathbf{e}}_3$ lie within and are perpendicular to each other. The local frame is orthonormal and independent of the local position within the cross-section. Concerning the deformed configuration we proceed in a similar way, but here the lateral slope vectors are, in general, no more perpendicular. Let $\mathbf{x} = \mathbf{x}(\xi, \eta, \zeta)$, then the local frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is given by

$$\mathbf{e}_1 = \frac{\mathbf{x}_{,\eta} \times \mathbf{x}_{,\zeta}}{\|\mathbf{x}_{,\eta} \times \mathbf{x}_{,\zeta}\|}, \quad \mathbf{e}_2 = \frac{\mathbf{x}_{,\zeta} \times \left(\mathbf{x}_{,\eta} \times \mathbf{x}_{,\zeta}\right)}{\|\mathbf{x}_{,\zeta} \times \left(\mathbf{x}_{,\eta} \times \mathbf{x}_{,\zeta}\right)\|}, \quad \mathbf{e}_3 = \frac{\mathbf{x}_{,\zeta}}{\|\mathbf{x}_{,\zeta}\|}. \tag{7}$$

Note, that the definition of the local frame is chosen arbitrarily, regarding its rotation about $\mathbf{e}_1$. Particularly, $\mathbf{x}_{,\zeta}$ defines the rotation of the cross-section around $\mathbf{e}_1$. Alternatively, $\mathbf{x}_{,\eta}$ could define the rotation of the cross-section, or a symmetric definition regarding the slope vectors could be built upon the polar decomposition—however, at higher computational costs.

With the local basis in the undeformed and the deformed configuration introduced, we can represent the respective rotation tensor from the reference to the undeformed configuration as

$$\bar{\mathbf{A}} = \bar{\mathbf{e}}_i \otimes \mathbf{e}_{\mathrm{ref},i}, \tag{8}$$

and that from the reference to the deformed configuration becomes

$$\mathbf{A}\bar{\mathbf{A}} = \mathbf{e}_i \otimes \mathbf{e}_{\mathrm{ref},i}. \tag{9}$$

2.1.1 Continuum Mechanics Formulation

Having provided the key ideas and assumptions concerning the geometry of deformation, the strain measures entering the constitutive equations are to be defined next. In the continuum mechanics formulation, Green's strain tensor is employed to measure the deformation,

$$\mathbf{E} = \frac{1}{2}\left(\mathbf{F}^T\mathbf{F} - \mathbf{I}\right), \tag{10}$$

where $\mathbf{F}$ denotes the deformation gradient, which is expressed as

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \bar{\mathbf{x}}} = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} \frac{\partial \boldsymbol{\xi}}{\partial \bar{\mathbf{x}}} = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} \left(\frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\xi}} \right)^{-1}, \tag{11}$$

since we want to use the referential coordinates for the sake of simplicity. Green's strain—the change of the metric represented in the reference configuration—is consequently given by

$$\mathbf{E} = \frac{1}{2} \left(\frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\xi}} \right)^{-T} \left\{ \left(\frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} \right)^{T} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} - \left(\frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\xi}} \right)^{T} \left(\frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\xi}} \right) \right\} \left(\frac{\partial \bar{\mathbf{x}}}{\partial \boldsymbol{\xi}} \right)^{-1}. \tag{12}$$

In case of an initially straight beam, the undeformed configuration is typically chosen as the reference configuration, i.e., $\boldsymbol{\xi} = \bar{\mathbf{x}}$. Accordingly, Green's strain tensor then reduces to the well-known representation

$$\mathbf{E} = \frac{1}{2} \left\{ \left(\frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} \right)^{T} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} - \mathbf{I} \right\}, \tag{13}$$

in which $\mathbf{I} = \mathbf{e}_{\mathrm{ref},i} \otimes \mathbf{e}_{\mathrm{ref},i}$ denotes the identity tensor.

While the choice of the above strain measure is natural within continuum theory, the structural mechanics formulation relies on the introduction of proper generalized strain measures which originate in the Cosserat theory of solids, which is described in what follows.

2.1.2 Structural Mechanics Formulation

Neglecting the in-plane deformation of cross-sections at first, a beam can be thought of as an elastic line with cross-sections attached to it. In the nonlinear rod model, the elastic line gets translated and stretched in the course of deformation; the cross-sections, which are represented by the local frames introduced above, undergo a rigid body rotation. The vector of generalized force strains describing both axial extension and shear deformation is the change of the derivatives of the axis's position vector with respect to the undeformed arc-length S:

$$\boldsymbol{\Gamma} = \frac{\partial \mathbf{r}}{\partial S} - \mathbf{A} \frac{\partial \bar{\mathbf{r}}}{\partial S}. \tag{14}$$

In order to compute the difference, the derivative in the deformed configuration is transformed into the local frame of the beam's undeformed configuration. Recalling that we want to express the involved field variables as functions of the referential coordinates, we use the relationship

$$dS = \sqrt{\left(\frac{\partial \bar{x}}{\partial \xi} \right)^{2} + \left(\frac{\partial \bar{y}}{\partial \xi} \right)^{2} + \left(\frac{\partial \bar{z}}{\partial \xi} \right)^{2}} d\xi = \left\| \frac{\partial \bar{\mathbf{r}}}{\partial \xi} \right\| d\xi \tag{15}$$

for rewriting the generalized force strains as

$$\boldsymbol{\Gamma} = \left\| \frac{\partial \bar{\mathbf{r}}}{\partial \xi} \right\|^{-1} \left(\frac{\partial \mathbf{r}}{\partial \xi} - \mathbf{A} \frac{\partial \bar{\mathbf{r}}}{\partial \xi} \right). \tag{16}$$

The definition of the generalized moment strains relies on the fundamental property of orthogonal tensors that

$$\mathbf{A}\mathbf{A}^T = \mathbf{I} \quad \Rightarrow \quad \frac{\partial \mathbf{A}}{\partial S}\mathbf{A}^T = -\mathbf{A}\frac{\partial \mathbf{A}^T}{\partial S} = -\left(\frac{\partial \mathbf{A}}{\partial S}\mathbf{A}^T \right)^T. \tag{17}$$

The vector of moment strains $\boldsymbol{\kappa}$ is the vector associated with the above skew-symmetric tensor such that the following identity holds for any vector $\mathbf{v}$,

$$\boldsymbol{\kappa} \times \mathbf{v} = \left(\frac{\partial \mathbf{A}}{\partial S}\mathbf{A}^T \right) \mathbf{v}. \tag{18}$$

For an alternative representation of the generalized moment strains, the vector of twist and curvature $\mathbf{k}$ is introduced,

$$\mathbf{k} = \frac{1}{2}\mathbf{e}_i \times \frac{\partial \mathbf{e}_i}{\partial S}, \tag{19}$$

which describes the change of the local basis along a material line,

$$\frac{\partial \mathbf{e}_i}{\partial S} = \mathbf{k} \times \mathbf{e}_i. \tag{20}$$

Likewise, the vector of the curvature and twist in the undeformed configuration is given by

$$\bar{\mathbf{k}} = \frac{1}{2}\bar{\mathbf{e}}_i \times \frac{\partial \bar{\mathbf{e}}_i}{\partial S}. \tag{21}$$

In terms of these vectors, the change of the local basis along the beam's axis can be written as

$$\frac{\partial \mathbf{A}}{\partial S} = \frac{\partial \mathbf{e}_i}{\partial S} \otimes \bar{\mathbf{e}}_i + \mathbf{e}_i \otimes \frac{\partial \bar{\mathbf{e}}_i}{\partial S} = (\mathbf{k} \times \mathbf{e}_i) \otimes \bar{\mathbf{e}}_i - \mathbf{e}_i \otimes (\bar{\mathbf{e}}_i \times \bar{\mathbf{k}}). \tag{22}$$

The product with the $\mathbf{A}^T$ yields

$$\frac{\partial \mathbf{A}}{\partial S}\mathbf{A}^T = (\mathbf{k} \times \mathbf{e}_i) \otimes \mathbf{e}_i - \mathbf{e}_i \otimes (\bar{\mathbf{e}}_i \times \bar{\mathbf{k}})\mathbf{A}^T = (\mathbf{k} \times \mathbf{e}_i) \otimes \mathbf{e}_i - \mathbf{e}_i \otimes (\mathbf{e}_i \times \mathbf{A}\bar{\mathbf{k}}), \tag{23}$$

where the identity $\mathbf{a} \times \mathbf{b} = (\mathbf{A}\mathbf{a} \times \mathbf{A}\mathbf{b})\mathbf{A}$ has been utilized. The skew-symmetry of the above tensor allows us to rewrite the product with some vector $\mathbf{v}$ as

$$\left(\frac{\partial \mathbf{A}}{\partial S}\mathbf{A}^T\right)\mathbf{v} = (\mathbf{k} \times \mathbf{v}) - (\mathbf{A}\bar{\mathbf{k}} \times \mathbf{v}) = (\mathbf{k} - \mathbf{A}\bar{\mathbf{k}}) \times \mathbf{v}. \tag{24}$$

Comparing this result with the previous definition (18), we can immediately identify the simple representation of $\boldsymbol{\kappa}$ in terms of $\mathbf{k}$ and $\bar{\mathbf{k}}$ as

$$\boldsymbol{\kappa} = \mathbf{k} - \mathbf{A}\bar{\mathbf{k}}. \tag{25}$$

2.2 Equations of Motion

Pursuing a finite element discretization, the equations of motion are discussed in their weak form. According to d'Alembert's principle in Lagrange's representation, the virtual work of the external forces is balanced by the sum of the virtual work of the internal forces, i.e., the variation of the strain energy, and the virtual work of the inertia forces

$$\delta W^{\text{inert}} + \delta W^{\text{int}} = \delta W^{\text{ext}}. \tag{26}$$

Similar to other beam formulations, the virtual work of external forces can be given in terms of products of concentrated forces and torques times virtual displacements and rotations, respectively. The virtual rotations need to be determined from the rotation tensor $\mathbf{A}$ and consequently from the slope vectors involved, cf. (6)–(7). The virtual work of surface tractions and body forces is obtained from surface and volume integrals over their products with the corresponding virtual displacements.

The virtual work of the inertia forces is given by the volume integral over the beam's domain Ω in the following reference configuration:

$$\delta W^{\text{inert}} = \int_{\Omega} \rho_0 \ddot{\mathbf{u}}^T \delta \mathbf{u} \, dV, \tag{27}$$

where the variation of the displacement field is indicated by a δ and ρ_0 denotes the referential density. Regardless of the particular kinematic hypothesis employed, the key idea of absolute displacements being interpolated results in a constant mass matrix. This property underlies all ANC elements discussed subsequently, apart from the ANC-like formulation concerning the thin spatial beam element with torsional stiffness of Sect. 3.5.

2.2.1 Continuum Mechanics Formulation

The virtual work of the internal forces in the continuum mechanics formulation corresponds to what is known from the conventional continuum theory of solids. Accordingly, the second Piola–Kirchhoff stress tensor $\mathbf{T}$ is work-conjugate to Green's strain tensor:

$$\delta W^{\text{int}} = \int_{\Omega} \mathbf{T} : \delta\mathbf{E}\, dV. \tag{28}$$

In case of a linearly elastic material—in finite strain theory, such constitutive behavior is referred to as St. Venant–Kirchhoff material—the stress tensor is given as

$$\mathbf{T} = {}^4\mathbf{D} : \mathbf{E}, \tag{29}$$

in which ${}^4\mathbf{D}$ is the fourth-order tensor of elastic moduli. In the isotropic case, e.g., it contains two independent parameters, i.e., Young's modulus E and Poisson's ratio ν. From a computational point of view, the distinction between vectors and tensors and their components with respect to some particular basis has to be taken care of at this point. In the numerical implementation, one often prefers to represent all quantities in the common inertial frame of the reference configuration $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$. When evaluating Green's strain tensor (12), the components are usually given with respect to the inertial frame whereas the components of the tensor of elastic moduli refer to the local frame in the undeformed configuration $\{\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3\}$. Therefore, one must either represent the components of ${}^4\mathbf{D}$ in the inertial frame when evaluating the stresses, or, alternatively, transform the components of the strain tensor into the local frame of the undeformed configuration using the rotation tensor $\bar{\mathbf{A}}$:

$$[\bar{\mathbf{E}}] = [\bar{\mathbf{A}}]^T [\mathbf{E}][\bar{\mathbf{A}}]. \tag{30}$$

In the above relation, we have introduced brackets in order to clearly distinguish between a tensor as an invariant object and its components relative to some tensorial basis. Subsequently, the components of the stress tensor can be determined from the constitutive equation (29), where a vector-matrix representation is typically employed for the sake of simplicity. Collecting the six independent components of both stress and strain tensor relative to the natural basis of the undeformed configuration $\{\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \bar{\mathbf{e}}_3\}$ in vectors,

$$\bar{\boldsymbol{\tau}} = \left[\bar{T}_{11}, \bar{T}_{22}, \bar{T}_{33}, \bar{T}_{23}, \bar{T}_{13}, \bar{T}_{12}\right]^T, \tag{31}$$

$$\bar{\varepsilon} = \left[\bar{E}_{11}, \bar{E}_{22}, \bar{E}_{33}, 2\bar{E}_{23}, 2\bar{E}_{13}, 2\bar{E}_{12}\right]^T, \tag{32}$$

Equation (29) can be equivalently rewritten in terms of the 6×6 matrix $\bar{\mathbf{D}}_{\text{CM}}$ as

$$\bar{\boldsymbol{\tau}} = \bar{\mathbf{D}}_{\text{CM}} \bar{\varepsilon}, \tag{33}$$

where $\bar{\mathbf{D}}_{\mathrm{CM}}$ gathers all relevant components of the fourth-order tensor ${}^4\mathbf{D}$. In case of a linearly elastic, isotropic material, for instance, the matrix is given by

$$\bar{\mathbf{D}}_{\mathrm{CM}} = \frac{E\nu^2}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & 1 & 1 & 0 & 0 & 0 \\ 1 & 1-\nu & 1 & 0 & 0 & 0 \\ 1 & 1 & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2\nu}{2}k_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2\nu}{2}k_3 \end{bmatrix}, \tag{34}$$

where k_2 and k_3 denote shear correction factors that account for the non-uniform distribution of shear stresses within the beam's cross-section. Note that these correction factors may be different from well-known structural mechanics shear correction factors due to the specific integration over the cross-section in ANC finite elements. While the shear correction factors provided above enable a correct transverse shear stiffness, a correction for torsional stiffness is not accounted for in the present continuum mechanics formulation.

The components of the second Piola–Kirchhoff stress tensor $\bar{\mathbf{T}}$, which are collected in the vector $\bar{\boldsymbol{\tau}}$, need to be transformed back into the reference frame afterwards

$$[\mathbf{T}] = [\bar{\mathbf{A}}][\bar{\mathbf{T}}][\bar{\mathbf{A}}]^T. \tag{35}$$

With the stress tensor given, the variation of Green's strain tensor remains to be determined when evaluating the virtual work of the internal forces (28):

$$\delta\mathbf{E} = \frac{1}{2}\left(\frac{\partial\bar{\mathbf{x}}}{\partial\xi}\right)^{-T} \left\{\left(\frac{\partial(\delta\mathbf{x})}{\partial\xi}\right)^T \frac{\partial\mathbf{x}}{\partial\xi} + \left(\frac{\partial\mathbf{x}}{\partial\xi}\right)^T \frac{\partial(\delta\mathbf{x})}{\partial\xi}\right\} \left(\frac{\partial\bar{\mathbf{x}}}{\partial\xi}\right)^{-1}. \tag{36}$$

2.2.2 Structural Mechanics Formulation

As opposed to the continuum mechanics formulation, the question of rational stress resultants that are conjugate to the previously introduced generalized strain measures is raised on a structural level. The internal forces and moments $\mathbf{f}$ and $\mathbf{m}$, respectively, represent stress resultants that can be regarded as quantities obtained upon a static condensation of the stress distribution within the cross-section relative to the beam's axis. The present variational formulation of the strain energy relies on the ideas of Reissner (1972, 1973), Antman (1972) and Simo (1985) according to which the internal forces are conjugate to the generalized force strains and the internal moments to the generalized moment strains, respectively,

$$\delta W^{\mathrm{int}} = \int_L \mathbf{f}^T \delta\boldsymbol{\Gamma} + \mathbf{m}^T \delta\boldsymbol{\kappa}\, d\xi, \tag{37}$$

where L denotes the length of the beam in the undeformed configuration. In the case of elastic material behavior, the constitutive equations for the cross-sectional forces and moments can be expressed as

$$\mathbf{f} = \mathbf{a}\boldsymbol{\Gamma} + \mathbf{c}^T\boldsymbol{\kappa}, \qquad \mathbf{m} = \mathbf{b}\boldsymbol{\kappa} + \mathbf{c}\boldsymbol{\Gamma}, \tag{38}$$

with $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ denoting second-order tensors of cross-sectional stiffnesses. Once again, the question of the respective basis of vectors and tensors involved needs to be addressed. The components of $\boldsymbol{\Gamma}$ and $\boldsymbol{\kappa}$ are typically available in the inertial frame that is used throughout the numerical analysis. The constitutive behavior (38), however, represents a locally linear behavior with a constant tangent stiffness that rotates with the beam in the course of deformation. Similar to the continuum mechanics approach, we have two options: one is determining the components of the material tensors relative to the inertial frame. Alternatively, the components of the generalized strains in the local frame of either the beam's undeformed or its deformed configuration are computed using the respective rotation tensor. Choosing the local frame in the undeformed configuration, the components of $\boldsymbol{\Gamma}$ and $\boldsymbol{\kappa}$ are transformed by means of

$$[\bar{\boldsymbol{\Gamma}}] = [\bar{\mathbf{A}}][\boldsymbol{\Gamma}], \qquad [\bar{\boldsymbol{\kappa}}] = [\bar{\mathbf{A}}][\boldsymbol{\kappa}]. \tag{39}$$

Again, we can gather the stress resultants and the generalized strains in vectors in order to represent the material behavior by means of a matrix equation:

$$\bar{\boldsymbol{\tau}}_{\mathrm{SM}} = \mathbf{D}_{\mathrm{SM}}\bar{\boldsymbol{\varepsilon}}_{\mathrm{SM}} \tag{40}$$

with

$$\bar{\boldsymbol{\tau}}_{\mathrm{SM}} = \left[\bar{f}_1, \bar{f}_2, \bar{f}_3, \bar{m}_1, \bar{m}_2, \bar{m}_3\right]^T, \quad \bar{\boldsymbol{\varepsilon}}_{\mathrm{SM}} = \left[\bar{\Gamma}_1, \bar{\Gamma}_2, \bar{\Gamma}_3, \bar{\kappa}_1, \bar{\kappa}_2, \bar{\kappa}_3\right]^T. \tag{41}$$

where $\mathbf{D}_{\mathrm{SM}}$ is the 6×6 cross-sectional stiffness matrix. In case of simple symmetric cross-sections, the coupling term disappears, i.e., $\mathbf{c} = \mathbf{0}$, and $\mathbf{D}_{\mathrm{SM}}$ becomes diagonal,

$$\mathbf{D}_{\mathrm{SM}} = \mathrm{diag}\,(EA, k_2GA_2, k_3GA_3, GJ_1, EI_2, EI_3)\,, \tag{42}$$

with commonly used beam properties, i.e., the axial stiffness EA, corrected shear stiffnesses $k_{2,3}GA_{2,3}$, torsional rigidity GJ_1 and bending stiffnesses $EI_{2,3}$.

To this point, the deformation of the cross-sections $\mathbf{u}_{\mathrm{cs}}$ in Eq. (4) has not been addressed within the structural mechanics formulation. In conventional beam theories, the cross-sections are usually assumed to be rigid, i.e., they only undergo a rotation relative to the undeformed configuration. Although such restriction has proven useful in many engineering applications, a significant change of the cross-sections size is inherent to certain problems as, e.g., rolling processes in metal processing. Among some of the ANC elements discussed subsequently, the parametrization facilitates including such deformation of the cross-sections' from a numerical point of view. For this purpose, the question of how to consistently augment the virtual work of the internal forces in terms of appropriate strain measures and conjugate forces needs to be answered. A natural approach is to extend the structural mechanics formulation by the corresponding terms in the continuum mechanics formulation. Following Eq. (4), the deformation gradient is expressed as

$$\mathbf{F} = \frac{\partial}{\partial \bar{\mathbf{x}}} \left\{ \mathbf{r} + \mathbf{A} \left(\bar{\mathbf{x}} - \bar{\mathbf{r}} \right) \right\} + \mathbf{G}_{\mathrm{cs}}, \tag{43}$$

where the displacement gradient $\mathbf{G}_{\mathrm{cs}}$ represents the additional contribution from the deformation of the cross-sections given by

$$\mathbf{G}_{\mathrm{cs}} = \frac{\partial \mathbf{u}_{\mathrm{cs}}}{\partial \bar{\mathbf{x}}}. \tag{44}$$

The definition of Green's strain (10) immediately reveals the coupling of the cross-sections' stretching and shearing with the conventional deformation allowed within Timoshenko's hypothesis. Subsequently, however, we introduce the key assumption that the in-plane deformation of the cross-sections does not interfere with the original structural mechanics formulation, or, in other words, the cross-section deformation is decoupled from generalized strain measures introduced above. Accordingly, only the in-plane components of the strain tensor (12), i.e.,

$$\bar{E}_{22} = \bar{\mathbf{e}}_2^T \left(\mathbf{E} \bar{\mathbf{e}}_2 \right), \quad \bar{E}_{33} = \bar{\mathbf{e}}_3^T \left(\mathbf{E} \bar{\mathbf{e}}_3 \right), \quad \bar{E}_{23} = \bar{\mathbf{e}}_3^T \left(\mathbf{E} \bar{\mathbf{e}}_2 \right), \tag{45}$$

are regarded when augmenting the virtual work of the internal forces. The above requirement further implies that the cross-sections' deformation does not affect the generalized forces and moments of the structural mechanics formulation, which—from a continuum mechanics perspective—represent cross-sectional resultants of the stresses. In case of an elastic material, for instance, we have to stipulate $\nu = 0$ such that the conjugate stresses are given by

$$T_{22} = E \bar{E}_{22}, \quad T_{33} = E \bar{E}_{33}, \quad T_{23} = 2G \bar{E}_{23}, \tag{46}$$

where E and G denote the Young's modulus and the shear modulus, respectively. The additional term in the virtual work of the internal forces consequently reads

$$\delta W_{\mathrm{cs}}^{\mathrm{int}} = \int_{\Omega} E \left(\bar{E}_{22} \delta \bar{E}_{22} + \bar{E}_{33} \delta \bar{E}_{33} \right) + 2G \bar{E}_{23} \delta \bar{E}_{23} dV. \tag{47}$$

If the in-plane strains are distributed uniformly within the cross-sections, the above relation simplifies to

$$\delta W_{\mathrm{cs}}^{\mathrm{int}} = \int_{L} EA \left(\bar{E}_{22} \delta \bar{E}_{22} + \bar{E}_{33} \delta \bar{E}_{33} \right) + 2\, GA\, \bar{E}_{23} \delta \bar{E}_{23} d\xi, \tag{48}$$

where the axial and shear stiffness have been introduced, which further connects the cross-sections' deformation to the structural mechanics formulation. The total variation of the internal forces is obtained by adding the contribution from the cross-sections (48) to the conventional expression for the virtual work of the internal forces (37):

$$\delta W^{\mathrm{int}}{}_{\mathrm{tot}} = \delta W^{\mathrm{int}} + \delta W^{\mathrm{int}}_{\mathrm{cs}}. \tag{49}$$

Before we proceed with the derivations, a few comments on the cross-sections' deformation seem to be appropriate. The numerous assumptions needed to eventually arrive at the simple expression (48) may appear restrictive to such an extent that the general applicability of the proposed formulation is questionable at best. The answer to that question is twofold: indeed but deliberately. The extension of the structural mechanics formulation for beams is not meant to contain all features of deformation a structure can be subjected to. It is specifically aimed at problems in which uniform in-plane stretch and shearing are relevant—as in the examples mentioned above—but the assumptions underlying the structural mechanics formulation are sufficient otherwise. That is to say, including the cross-sectional deformation widens the scope of applicability of the efficient structural mechanics formulation. In problems showing a more complex state of deformation, for which the coupling of in-plane and out-of-plane deformation cannot be neglected, the continuum mechanics formulation needs to be resorted to.

From a numerical point of view, the expressions to be evaluated in the general case of a beam that is arbitrarily curved in its undeformed configuration are relatively complicated since both the generalized strains and conjugate forces of the structural mechanics approach and the components of the strain tensor and the conjugate stresses of the continuum mechanics formulation are required. For an initially straight beam, however, the terms related to the cross-sectional deformation simplify significantly. In this case, the relevant components of Green's strain tensor with respect to the global frame are given by

$$E_{\eta\eta} = \frac{1}{2}\left(\frac{\partial \mathbf{x}}{\partial \eta}\frac{\partial \mathbf{x}}{\partial \eta} - 1\right), \; E_{\zeta\zeta} = \frac{1}{2}\left(\frac{\partial \mathbf{x}}{\partial \zeta}\frac{\partial \mathbf{x}}{\partial \zeta} - 1\right), \; E_{\eta\zeta} = \frac{1}{2}\frac{\partial \mathbf{x}}{\partial \eta}\frac{\partial \mathbf{x}}{\partial \zeta}. \tag{50}$$

Some of the ANC elements discussed subsequently are based on the interpolation of the derivatives contained in the above relations which greatly facilitates the evaluation of the strains related to the cross-sectional deformation.

2.3 Numerical Interpolation

The fundamental idea of ANCF is the direct interpolation of positions and position gradients with respect to the global frame—therefore, absolute—using positions and position gradients of a finite number of points, i.e., the nodes. Accordingly, the position vector—or rather, its components with respect to the global frame—of a beam's material point is represented by a Ritz approach as

$$\mathbf{x}(\boldsymbol{\xi}, t) = \mathbf{S}(\boldsymbol{\xi})\mathbf{q}(t), \quad \mathbf{x} \in \mathbb{R}^m \tag{51}$$

where $\mathbf{q}$ denotes the vector of n generalized coordinates and $\mathbf{S}$ is the $m \times n$ matrix of interpolation or shape functions, which is briefly referred to as shape function matrix. Naturally, the same representation is used for the position field in the undeformed configuration,

$$\bar{\mathbf{x}}(\boldsymbol{\xi}) = \mathbf{S}(\boldsymbol{\xi})\bar{\mathbf{q}}. \tag{52}$$

Regarding both formulation and implementation, it should be mentioned that it is more or less a matter of taste of whether absolute nodal positions or displacements are utilized as generalized coordinates. Employing a Galerkin projection, the variation of the position is contained in the same function space as the position vector itself, i.e., we use the same shape function matrix

$$\delta\mathbf{x}(\boldsymbol{\xi}) = \mathbf{S}(\boldsymbol{\xi})\delta\mathbf{q}. \tag{53}$$

2.4 Overview of Different ANC Finite Elements

In the absolute nodal coordinate formulation, the design of finite elements is based on the choice of nodal degrees of freedom (coordinates).

In most ANC finite elements, the nodal coordinates consist of position or displacement coordinates as well as the corresponding derivatives with respect to the referential coordinates (ξ, η, ζ).

Figure 2 shows selected 2D and 3D ANC finite elements. As a minimum, one axial slope vector is employed in order to create a Bernoulli–Euler beam finite elements, see Fig. 2a, b. Another case is retrieved, if all components of the gradient at each node are used to define shear and cross-section deformable ANC finite elements, also denoted as fully parametrized, see Fig. 2c, d. The term 'fully parametrized' is used, because all components of the gradient at the nodal positions are parametrized by three nodal slope vectors.

The coordinates of two- and three-noded beam finite elements according to Eq. (51) can be given in the general form,

$$\begin{aligned} \mathbf{q}^{(2\,\mathrm{node})} &= \left[\mathbf{q}^{(1)^T} \quad \mathbf{q}^{(2)^T}\right]^T, \quad \text{and} \\ \mathbf{q}^{(3\,\mathrm{node})} &= \left[\mathbf{q}^{(1)^T} \quad \mathbf{q}^{(2)^T} \quad \mathbf{q}^{(3)^T}\right]^T. \end{aligned} \tag{54}$$

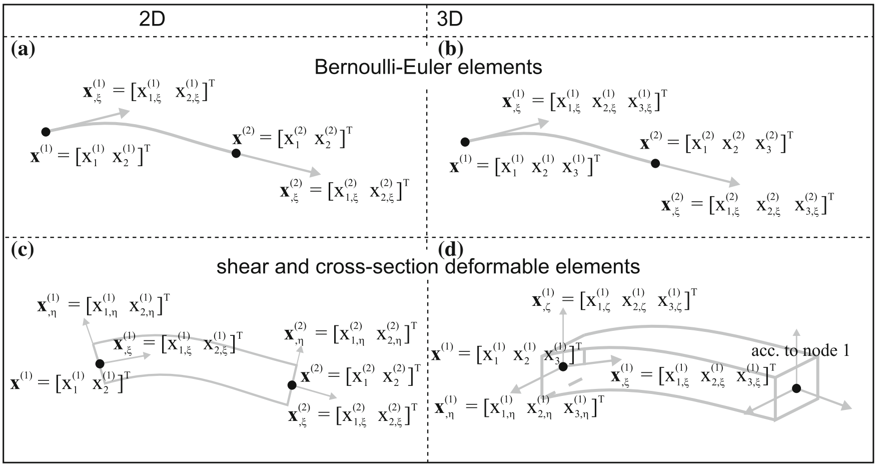

2D
3D
(a)
(b)
Bernoulli-Euler elements
$\mathbf{x}_{,\xi}^{(1)} = [x_{1,\xi}^{(1)} \; x_{2,\xi}^{(1)}]^T$
$\mathbf{x}^{(1)} = [x_1^{(1)} \; x_2^{(1)}]^T$
$\mathbf{x}^{(2)} = [x_1^{(2)} \; x_2^{(2)}]^T$
$\mathbf{x}_{,\xi}^{(2)} = [x_{1,\xi}^{(2)} \; x_{2,\xi}^{(2)}]^T$
$\mathbf{x}_{,\xi}^{(1)} = [x_{1,\xi}^{(1)} \; x_{2,\xi}^{(1)} \; x_{3,\xi}^{(1)}]^T$
$\mathbf{x}^{(1)} = [x_1^{(1)} \; x_2^{(1)} \; x_3^{(1)}]^T$
$\mathbf{x}^{(2)} = [x_1^{(2)} \; x_2^{(2)} \; x_3^{(2)}]^T$
$\mathbf{x}_{,\xi}^{(2)} = [x_{1,\xi}^{(2)} \; x_{2,\xi}^{(2)} \; x_{3,\xi}^{(2)}]^T$
(c)
(d)
shear and cross-section deformable elements
$\mathbf{x}_{,\eta}^{(1)} = [x_{1,\eta}^{(1)} \; x_{2,\eta}^{(1)}]^T$
$\mathbf{x}_{,\xi}^{(1)} = [x_{1,\xi}^{(1)} \; x_{2,\xi}^{(1)}]^T$
$\mathbf{x}^{(1)} = [x_1^{(1)} \; x_2^{(1)}]^T$
$\mathbf{x}_{,\eta}^{(2)} = [x_{1,\eta}^{(2)} \; x_{2,\eta}^{(2)}]^T$
$\mathbf{x}^{(2)} = [x_1^{(2)} \; x_2^{(2)}]^T$
$\mathbf{x}_{,\xi}^{(2)} = [x_{1,\xi}^{(2)} \; x_{2,\xi}^{(2)}]^T$
$\mathbf{x}_{,\zeta}^{(1)} = [x_{1,\zeta}^{(1)} \; x_{2,\zeta}^{(1)} \; x_{3,\zeta}^{(1)}]^T$
$\mathbf{x}^{(1)} = [x_1^{(1)} \; x_2^{(1)} \; x_3^{(1)}]^T$
acc. to node 1
$\mathbf{x}_{,\xi}^{(1)} = [x_{1,\xi}^{(1)} \; x_{2,\xi}^{(1)} \; x_{3,\xi}^{(1)}]^T$
$\mathbf{x}_{,\eta}^{(1)} = [x_{1,\eta}^{(1)} \; x_{2,\eta}^{(1)} \; x_{3,\eta}^{(1)}]^T$

Fig. 2 Overview of some basic ANC finite elements. **a** 8 DOF, planar ANC finite element, **b** 12 DOF, spatial ANC finite element, **c** 12 DOF, planar ANC finite element with shear and cross-section deformation, **d** 24 DOF, spatial ANC finite element with shear and cross-section deformation

in which $\mathbf{q}^{(i)^T}$ represents the nodal coordinates of the i-th node. Following the original idea of the ANCF, a fully parametrized set of nodal position and slope vectors has been utilized,

$$\mathbf{q}_{\text{fp}}^{(j)} = \left[\mathbf{x}^{(j)^T} \quad \mathbf{x}_{,\xi}^{(j)^T} \quad \mathbf{x}_{,\eta}^{(j)^T} \quad \mathbf{x}_{,\zeta}^{(j)^T}\right]^T . \tag{55}$$

The vector $\mathbf{x}^{(j)}$ represents the current position of the node j of the beam finite element. Note that the nodal coordinates of Eq. (55) are comprised of the position and three slope vectors which represent the deformation gradient.

In order to efficiently model ANC beam finite elements based on the Bernoulli–Euler theory, so-called gradient-deficient nodal coordinates are utilized, which means that not all components of the gradient are employed in the nodal coordinates,

$$\mathbf{q}_{\text{axial}}^{(j)} = \left[\mathbf{x}^{(j)^T} \quad \mathbf{x}_{,\xi}^{(j)^T}\right]^T . \tag{56}$$

In case of ANC beam finite elements which cover the Timoshenko beam theory, gradient-deficient nodal coordinate that which do not contain the axial slope vector are frequently used

$$\mathbf{q}_{\text{cross-section}}^{(j)} = \left[\mathbf{x}^{(j)^T} \quad \mathbf{x}_{,\eta}^{(j)^T} \quad \mathbf{x}_{,\zeta}^{(j)^T}\right]^T . \tag{57}$$

3 ANC Finite Elements Based on the Bernoulli-Euler Condition

In this section, the 2D and 3D formulations of thin beam (or cable) finite elements based on the ANCF are discussed. The original formulations of 2D Bernoulli–Euler ANC beam finite elements have been developed by Shabana and Schwertassek (1997) and later on by Berzeri and Shabana (2002). In the present section, an extended formulation is presented for 2D and 3D thin beams, which follows the works of Gerstmayr and Shabana (2006), Gerstmayr and Irschik (2008) and Gruber et al. (2013).

3.1 Kinematics of Thin ANC Beam Finite Elements

For notational convenience, the derivative of a quantity with respect to the axial coordinate ξ is subsequently abbreviated as

$$\frac{\partial\,()}{\partial \xi} = ()' \,. \tag{58}$$

The two-noded planar element has eight degrees of freedom, see Fig. 2a. For such beam element, the position (or displacement) of its axis can be interpolated by two third-order polynomials in ξ,

$$\mathbf{x}^{2D} = \begin{bmatrix} x_1^{2D} \\ x_2^{2D} \end{bmatrix} = \begin{bmatrix} a_0 + a_1\xi + a_2\xi^2 + a_3\xi^3 \\ b_0 + b_1\xi + b_2\xi^2 + b_3\xi^3 \end{bmatrix}. \tag{59}$$

The coefficients a_i and b_i are determined by requiring that the generalized degrees of freedom $\mathbf{q}^{2D}$ represent components of the nodal positions (or displacements) and slope vectors. Using third-order polynomials also for the interpolation of the slope vectors, we obtain the shape functions S_i,

$$\begin{aligned} S_1 &= \frac{1}{2} - \frac{3}{4}\xi + \frac{1}{4}\xi^3, \quad S_2 = \frac{L}{8}\left(1 - \xi - \xi^2 + \xi^3\right), \\ S_3 &= \frac{1}{2} + \frac{3}{4}\xi - \frac{1}{4}\xi^3, \quad S_4 = \frac{L}{8}\left(-1 - \xi + \xi^2 + \xi^3\right). \end{aligned}$$

which are gathered in the shape function matrix $\mathbf{S}_m$ as

$$\mathbf{x}^{2D} = [S_1\mathbf{I} \quad S_2\mathbf{I} \quad S_3\mathbf{I} \quad S_4\mathbf{I}]\mathbf{q}^{2D} = \mathbf{S}_m\mathbf{q}^{2D}, \tag{60}$$

in which $\mathbf{I}^{2D}$ is the 2×2 unit matrix.

In addition to the thin planar ANC beam element, two formulations for spatial (3D) thin ANC finite elements exist. The simplest spatial element considers bending and axial stretch only, see Gerstmayr and Shabana (2006), and thus can only be used to model cable problems, whereas an extended formulation for spatial beam elements can also handle torsion. The latter extends the idea of Dmitrochenko and Pogorelov (2003) in order to prevent from singularities, see Gruber et al. (2013) or Sect. 3.5.

For thin spatial beams, the polynomial interpolation of the position reads

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_0 + a_1\xi + a_2\xi^2 + a_3\xi^3 \\ b_0 + b_1\xi + b_2\xi^2 + b_3\xi^3 \\ c_0 + c_1\xi + c_2\xi^2 + c_3\xi^3 \end{bmatrix}. \tag{61}$$

Again, the coefficients a_i, b_i and c_i are chosen such that the generalized coordinates $\mathbf{q}$ correspond to the components of the position (or displacement) and slope vectors at the nodes. For a compact representation of the relation between the position vector and the element coordinates, the shape functions can be collected in the shape function matrix,

$$\mathbf{x} = [S_1\mathbf{I} \quad S_2\mathbf{I} \quad S_3\mathbf{I} \quad S_4\mathbf{I}]\mathbf{q} = \mathbf{S}_m\mathbf{q}. \tag{62}$$

Note that we will not distinguish between planar (^{2D}) and spatial vectors in the following since most mathematical operations are identical. In the planar case, the vector product ($\times$) is understood as the product of two spatial vectors that represent the embedding of the planar ones in the 3D space.

3.2 Virtual Work of Elastic Forces for Thin Beams Without Torsional Stiffness

In thin ANC beam finite elements, only a structural mechanics formulation exists for the definition of the elastic forces, while in the thick ANC beam finite elements, both a continuum mechanics and structural mechanics formulations are available for the computation of the elastic forces.

3.2.1 Bending and Axial Strain

In the present section, the kinematics and the strain energy of a planar Bernoulli–Euler beam undergoing large rigid body motions and large deformations (but small strains) is investigated. In order to keep this section simple, the planar beam formulation is written for an initially straight and undeformed beam, assuming that the undeformed configuration is identical to the reference configuration (beam aligned along $\mathbf{e}_x$ axis).

The kinematics of the beam element is according to Fig. 1. In a planar Bernoulli–Euler beam, Eq. (4) reduces to

$$\mathbf{x}(\xi, \eta) = \mathbf{r}(\xi) + \eta\,\mathbf{e}_2(\xi). \tag{63}$$

The local basis, which is rigidly attached to the cross-section of the beam in current configuration, is simply defined within the relations

$$\mathbf{e}_1 = \frac{1}{\|\mathbf{r}'\|}\frac{\partial \mathbf{r}}{\partial \xi}, \quad \mathbf{e}_1^T\mathbf{e}_2 = 0, \quad \mathbf{e}_2^T\mathbf{e}_z = 0, \quad \text{and} \quad \mathbf{e}_3 = \mathbf{e}_z. \tag{64}$$

The derivative of the position vector $\mathbf{x}$ with respect to ξ is given by

$$\mathbf{x}' = \frac{\partial \mathbf{x}}{\partial \xi} = \mathbf{r}'(\xi) + \eta\,\mathbf{e}_2'(\xi). \tag{65}$$

The derivative of the cross-section vector $\mathbf{e}_2$ with respect to ξ follows as

$$\mathbf{e}_2' = -\theta'\mathbf{e}_1. \tag{66}$$

Thus, the rate of change of the rotation of the cross-section $\partial\theta/\partial S$, also denoted as material measure of curvature K, is given by

$$K = \frac{\partial \theta}{\partial S} = \frac{\partial \theta}{\partial \xi}\frac{\partial \xi}{\partial S} = \frac{1}{\|\bar{\mathbf{r}}'\|}\left(\frac{\mathbf{r}' \times \mathbf{r}''}{\|\mathbf{r}'\|^2}\right)^T \mathbf{e}_3. \tag{67}$$

The latter result follows from the general definition of the moment strain measure (25). In the planar case, the only nontrivial component of the vector of twist and curvature $\mathbf{k}$ reads

$$\mathbf{k}^T\mathbf{e}_z = \frac{1}{2}\left(\mathbf{e}_1 \times \frac{\partial \mathbf{e}_1}{\partial S} + \mathbf{e}_2 \times \frac{\partial \mathbf{e}_2}{\partial S}\right)^T \mathbf{e}_z = \left(\mathbf{e}_1 \times \frac{\partial \mathbf{e}_1}{\partial S}\right)^T \mathbf{e}_z, \tag{68}$$

where the identity $\mathbf{e}_2 = \mathbf{e}_z \times \mathbf{e}_1$ has been utilized. Introducing Eq. (64) and using the relation (15), the above equation yields

$$\mathbf{k}^T\mathbf{e}_z = \left(\frac{\mathbf{r}'}{\|\mathbf{r}'\|} \times \frac{\mathbf{r}''}{\|\mathbf{r}'\|}\frac{1}{\|\bar{\mathbf{r}}'\|}\right)^T \mathbf{e}_z = \frac{1}{\|\bar{\mathbf{r}}'\|}\left(\frac{\mathbf{r}' \times \mathbf{r}''}{\|\mathbf{r}'\|^2}\right)^T \mathbf{e}_z = K. \tag{69}$$

Assuming that the beam's axis may be curved but not stretched in the undeformed configuration, i.e., $\|\bar{\mathbf{r}}'\| = 1$, we obtain the familiar relation

$$K = \left(\frac{\mathbf{r}' \times \mathbf{r}''}{\|\mathbf{r}'\|^2}\right)^T \mathbf{e}_3. \tag{70}$$

Finally, the derivative of the position vector $\mathbf{x}$ reads

$$\mathbf{x}' = \left(\|\mathbf{r}'(\xi)\| - \eta K\right)\mathbf{e}_1. \tag{71}$$

Thus, the computation of the deformation gradient simply becomes

$$\begin{aligned}\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \xi} \otimes \mathbf{e}_1 + \frac{\partial \mathbf{x}}{\partial \eta} \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3 \\ = \left(\left\|\mathbf{r}'(\xi)\right\| + \eta K\right) \mathbf{e}_1 \otimes \mathbf{e}_x + \mathbf{e}_2 \otimes \mathbf{e}_y + \mathbf{e}_3 \otimes \mathbf{e}_z.\end{aligned} \tag{72}$$

Note that the condition $\mathbf{e}_3 = \mathbf{e}_z$ holds in the planar case. It immediately follows that the only nonzero component of the Green strain tensor,

$$\mathbf{E} = \frac{1}{2}(\mathbf{F}^T\mathbf{F} - \mathbf{I}), \tag{73}$$

in the local frame $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is given as

$$E_{11} = \frac{1}{2}\left\{\left(\|\mathbf{r}'\| + \eta K\right)^2 - 1\right\}. \tag{74}$$

Usually, Green's strain tensor is not used in beam theories. Its quadratic dependency on the beam's cross-section coordinate η leads to nonzero strain at the beam axis for pure bending, see Gerstmayr and Irschik (2008).

Therefore, the strain components are usually linearized with respect to the cross-section coordinates. In the planar case of Bernoulli–Euler beams, a more elegant way to obtain geometrically linearized strain measures is shown subsequently. Biot's strain tensor is obtained from the polar decomposition of the deformation gradient,

$$\mathbf{F} = \mathbf{R}\mathbf{U}, \tag{75}$$

in which $\mathbf{R}$ denotes the rotational part of the deformation gradient and $\mathbf{U}$ represents the stretch, which is related to Biot's strain by $\mathbf{H} = \mathbf{U} - \mathbf{I}$. Due to the simple structure of the deformation gradient in the planar case, it follows that $\mathbf{R} = \mathbf{A}$, which results in

$$\mathbf{U} = \mathbf{A}^T\mathbf{F} = \left(\|\mathbf{r}'\| - \eta K\right) \mathbf{e}_1 \otimes \mathbf{e}_x + \mathbf{e}_2 \otimes \mathbf{e}_y + \mathbf{e}_3 \otimes \mathbf{e}_z. \tag{76}$$

The work-conjugate stress to the Biot strain $\mathbf{H}$ is the Biot stress $\mathbf{B}$. Under the assumption of a linear elastic material, the following relation can be applied:

$$B_{11} = EH_{11}, \tag{77}$$

in which E represents the Young's modulus.

In the beam theory, the strain component $H_{11} = \varepsilon_0 + \varepsilon_{bend}$ is split into a mean value, the (sectional) axial strain ε_0 and the bending strain proportional to the curvature K,

$$\varepsilon_0 = \|\mathbf{r}'\| - 1 \quad \text{and} \quad \varepsilon_{bend} = \eta K. \tag{78}$$

Finally, the stress resultants are introduced as the normal force

$$N = \int_A B_{11} dA = \int_A E(\varepsilon_0 - \eta K) dA = EA\varepsilon_0, \tag{79}$$

and the bending moment

$$M = \int_A \eta \bar{B}_{11} dA = \int_A E(\eta\,\varepsilon_0 - \eta^2 K) dA = EI\,K, \tag{80}$$

where the beam's axis is chosen such that

$$\int_A E\eta dA = 0. \tag{81}$$

In order to consider curved and pre-stretched beams, the curvature $\bar{K}$ and stretch $\bar{\varepsilon}_0$ in the undeformed configuration need to be considered,

$$N = EA(\varepsilon_0 - \bar{\varepsilon}_0), \quad M = EI\,(K - \bar{K}). \tag{82}$$

The relations for the sectional strain measures (78) as well as the stress resultants (82) represent the conventional linear elastic beam modeling for large deformation beams, which has been used, e.g., for the extensible Euler elastica, Reissner's shear deformable beam Reissner (1972) or the geometrically exact beam model of Simo and Vu-Quoc (1986a).

The virtual work of elastic forces for the sectional strain measures and stress resultants based on Biot's strain is provided as

$$\delta W_S = \int_L N\delta\varepsilon_0 - M\delta K\, d\xi. \tag{83}$$

In contrast, the St. Venant–Kirchhoff material model (29) can be used instead. The sectional strain measures as well as the stress resultants can be computed in a similar fashion from Eq. (74). For details of the derivation of the stress resultants, see Gerstmayr and Irschik (2008) and Irschik and Gerstmayr (2009b) for shear-deformable beams. The stress resultants for the St. Venant–Kirchhoff material model can be computed from the first Piola–Kirchhoff stress tensor, see Appendix A of Gerstmayr and Irschik (2008), and result in

$$N^{(P1)} = \varepsilon_{11}^0 \|\mathbf{r}'\| + \frac{3}{2} EIK^2 \|\mathbf{r}'\|, \tag{84}$$

and

$$M^{(P1)} = -EI\,K\|\mathbf{r}'\|^2 + \frac{1}{2} EI_4 K^3. \tag{85}$$

Obviously, the fourth area moment of inertia EI_4 enters the bending moment due to the nonlinear distribution along the cross-section of the Green–Lagrange strain, see Fig. 2 of Gerstmayr and Irschik (2008). Equations (84)–(85) provide insight into what happens in a continuum mechanics based formulation of an ANC beam finite element, which is usually based on the St. Venant–Kirchhoff material.

The virtual work of elastic forces results into the classical form

$$\delta W_S^{SVK} = \int_L N^{(P1)} \delta\varepsilon_0^{(G)} + M^{(P1)} \delta K^{(G)} \, d\xi, \tag{86}$$

taking into account the cross-sectional strain measures based on Green's strain, which is indicated by a superscript '(G)',

$$\varepsilon_0^{(G)} = \varepsilon_{11}^0 + \frac{1}{2}\frac{EI}{EA}K^2, \tag{87}$$

and

$$K^{(G)} = K\|\mathbf{r}'\|. \tag{88}$$

The quadratic dependency of the axial strain $\varepsilon_0^{(G)}$ on the square of the curvature K can be explained in terms of the quadratic distribution of Green's strains, see Fig. 2 of Gerstmayr and Irschik (2008).

A comparison of Eqs. (83) and (86) reveals the difference of a continuum mechanics and a structural mechanics model of a Bernoulli–Euler beam in the ANCF. This idea can be extended to 3D and shear-deformable beams, as well. The most important contribution, however, is due to axial strain and bending.

3.3 Linearized Axial and Bending Strain and Relation to Floating Frame of Reference Formulation

The Biot's strain component (78) corresponds to a linearization of the local strain components with respect to the local frame of the cross-section.

In the case of the Biot's strain and Bernoulli–Euler beam theory, the polar decomposition exactly gives the rotation of the cross-section as the rotational part of the deformation gradient, cf. Eq. (76). In order to further simplify the beam finite element, it is possible to use a linearization about an average rotation of the whole beam element. Early development of the ANCF, see Shabana and Schwertassek (1997) and Escalona et al. (1998), discussed the stiffness matrix of the ANC finite element for such element-wise linearization.

A planar co-rotational coordinate system **i** and **j** has been introduced,

$$\mathbf{i} = \frac{\mathbf{r}^{(2)} - \mathbf{r}^{(1)}}{\|\mathbf{r}^{(2)} - \mathbf{r}^{(1)}\|}, \tag{89}$$

in which $\mathbf{r}^{(1)}$, $\mathbf{r}^{(2)}$ are the positions of the left and the right node of the finite element. The second local axis is perpendicular to $\mathbf{i}$, i.e.,

$$\mathbf{j} = \begin{bmatrix} -i_2 \\ i_1 \end{bmatrix}, \tag{90}$$

In this way, the vector $\mathbf{u}$ is introduced

$$\mathbf{u} = \mathbf{r}(\xi) - \mathbf{r}^{(1)}, \tag{91}$$

and the projection of $\mathbf{u}$ into the local element frame leads to the relations for the local beam deformation quantities

$$\mathbf{u}_d = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \mathbf{u}^T\mathbf{i} - \xi \\ \mathbf{u}^T\mathbf{j} \end{bmatrix}. \tag{92}$$

Thus, the strain energy can be written as

$$U = \frac{1}{2}\int_L EAu'^2 + EIv''^2\, d\xi = \frac{1}{2}\int_L EA\left\{\left(\mathbf{u}'\right)^T\mathbf{i} - 1\right\}^2 + EI\left\{\left(\mathbf{u}''\right)^T\mathbf{j}\right\}^2 d\xi. \tag{93}$$

The latter approach fully corresponds to the floating frame of reference formulation, which assumes geometrically linearized relations in the body or element frame, see Shabana and Schwertassek (1997). An extension of this idea to shear-deformable 3D ANC beams has been introduced by Gerstmayr (2009), in which the linearized strains are computed in a co-rotational configuration of the large deformation beam element.

In a further work, a comparison of the floating frame of reference formulation based on geometrically linearized relations in each beam finite element to the ANC beam finite element with fully geometrically nonlinear formulation has been performed by Dibold et al. (2009). It turned out that co-rotationally linearized finite elements converge to exactly the same solution of large deformation static and dynamics examples as compared to Bernoulli–Euler ANC beam finite elements as discussed in the present section. The CPU performance of both formulations is similar and mainly depends on the type of mechanical problem.

3.4 Thin 3D ANC Beam Finite Element Without Torsional Stiffness

The planar Bernoulli–Euler ANC beam finite element can be extended to 3D straightforward, by adding a third component to the position and slope degrees of freedom, see Gerstmayr and Shabana (2006). In this way, a specific cable finite element is found, which has the following restrictions:

(a) The bending stiffness must be symmetrical, $EI_{\eta\eta} = EI_{\zeta\zeta}$, which applies to homogeneous round or quadratic cables
(b) The torsional stiffness and the moment of inertia for a rotation of the cross-section about the beam's axis is neglected; thus, it must be guaranteed that the physical problem which is modeled with the beam finite element does not show a twist or rotation about the beam's axis

If these restrictions are fulfilled, the 3D cable finite element becomes extremely simple. The axial strain is identical to the planar case (78),

$$\varepsilon_0 = \|\mathbf{r}'\| - 1, \tag{94}$$

the bending strain (material measure of curvature) is derived from Eq. (18). Due to simplicity of the structure of the rotation matrix, the formula of the curvature reads

$$K = \frac{\|\mathbf{r}' \times \mathbf{r}''\|}{\|\mathbf{r}'\|^2}. \tag{95}$$

Note that in the original work of Gerstmayr and Shabana (2006), slightly different strain measures have been used, which has been discussed and corrected in the work of Gerstmayr and Irschik (2008).

The virtual work of elastic forces for the ANC cable finite element is given by Eqs. (37) and (40), using only the axial stiffness and the bending stiffness.

The ANC cable finite element is superior to other finite elements because of its simple structure and the resulting computational efficiency. If torsion plays an important role, however, then it needs to be extended as described in the following section.

3.5 *Thin 3D ANC Beam Finite Element with Torsional Stiffness*

If torsional deformation is considered additionally to bending and axial deformation of a spatial beam, then the correct representation of its configuration in space requires additional information addressing the rotation of the cross-section about the beam axis (at every point of the beam axis).

Particularly, let us choose a fixed ξ for which $\bar{\mathbf{r}}(\xi)$ and $\mathbf{r}(\xi)$ denote the position of an axial point in reference and actual configuration, respectively (see Fig. 1). The straightforward way in the ANCF to describe the rotation of a beam's cross-section at this particular point would be to consider three more absolute nodal coordinates in form of a slope vector in lateral direction, see Yakoub and Shabana (2001). However, this vector would have to yield two more conditions: first, being perpendicular to the beams axis (which is one of the basic assumptions in Bernoulli–Euler beam theory), and second, remaining its length constant in order to avoid thickness deformation of

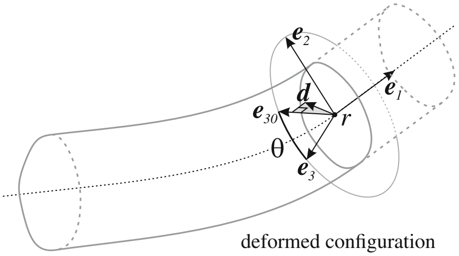


Fig. 3 Geometrical description of a thin beam with torsional stiffness. The orientation of the cross-section at point r is defined by the normalized projection $\mathbf{e}_{30}$ of the director $\mathbf{d}$ into the normal plane of the beam axis, and a subsequent rotation about the beam axis by the torsional angle θ, which gives $\mathbf{e}_3$

the beam. Owing to that, only one degree of freedom remains to be chosen (addressing the torsional rotation of the lateral slope vector about the beam axis) in order to fully describe the beam's configuration.

We express this single degree of freedom by the natural choice of a torsional angle $\theta(\xi)$ (see Fig. 3). Note, that this angle is no more an absolute, but a relative quantity. It is measured relative to the projection of a vector $\mathbf{d}(\xi)$, called director, into the normal plane of the axial slope $\mathbf{r}'(\xi)$. The torsional angle of the cross-section in reference configuration is denoted by $\bar{\theta}(\xi)$ and measured also relative to the orientation of the director $\mathbf{d}(\xi)$, i.e., its projection into the normal plane of $\bar{\mathbf{r}}'(\xi)$. Note that the director $\mathbf{d}(\xi)$, other than the axial position or slope vectors, basically represents a constant vector in time t. Let us—for the moment, and the sake of simplicity—additionally assume, that the director is constant in space, meaning $\mathbf{d}(\xi) = \mathbf{d}$ for all ξ, and omit the explicit notion of the variables ξ and t in the following formulas. The rotation of the local frame at a particular axis point, see Eq. (6), may be defined as

$$\mathbf{e}_1 = \frac{\mathbf{r}'}{|\mathbf{r}'|}\,, \tag{96}$$

$$\mathbf{e}_2 = \mathbf{e}_{20}\cos(\theta) + \mathbf{e}_{30}\sin(\theta)\,, \tag{97}$$

$$\mathbf{e}_3 = \mathbf{e}_{30}\cos(\theta) - \mathbf{e}_{20}\sin(\theta)\,, \tag{98}$$

in which $\mathbf{e}_{30}$ denotes the normalized projection of the director $\mathbf{d}$ into the normal plane of the axial slope $\mathbf{r}'$, i.e.,

$$\mathbf{e}_{30} = \frac{\hat{\mathbf{e}}_{30}}{\left|\hat{\mathbf{e}}_{30}\right|}\,, \quad \hat{\mathbf{e}}_{30} = \mathbf{d} - (\mathbf{d}^T\,\mathbf{e}_1)\,\mathbf{e}_1\,, \tag{99}$$

and the bi-normal $\mathbf{e}_{20}$ is obtained by the cross product

$$\mathbf{e}_{20} = \mathbf{e}_{30} \times \mathbf{e}_1\,. \tag{100}$$

Thereby, the curvature strain $\boldsymbol{\kappa}$ from Eq. (18) and the axial strain $\varepsilon_0 = \boldsymbol{\Gamma}\,\mathbf{e}_x$, utilizing $\boldsymbol{\Gamma}$ from Eq. (16), can be computed. Combining these strain measures together with the assumption of vanishing shear strains, i.e.,

$$|\boldsymbol{\Gamma}^T\,\mathbf{e}_y| + |\boldsymbol{\Gamma}^T\,\mathbf{e}_z| = 0\,,$$

the variational formulation of the strain energy according to Eq. (37) is fully defined. Note that a spatially non-constant director approach may be required combined with a temporal director update (see Sect. 3.6) in order to guarantee the Gram–Schmidt projection in Eq. (99) being well defined along the beam's axis.

Let us turn to the spatial discretization by means of an ANC beam finite element with two nodes. In addition to the cubic interpolation of the axial position $\mathbf{r}$, see Eq. (62), also the torsional angle at ξ is obtained by interpolation between nodal degrees of freedom,

$$\theta(\xi) = \mathbf{S}_{m\theta}(\xi)\mathbf{q}_\theta\,. \tag{101}$$

The presented ANC beam finite element provides a linear interpolation of the torsional angle

$$\mathbf{S}_{m\theta}(\xi) = [S_5(\xi) \quad S_6(\xi)]\,, \quad S_5(\xi) = \frac{1}{2} - \frac{\xi}{L}\,, \quad S_6(\xi) = \frac{1}{2} + \frac{\xi}{L}\,, \tag{102}$$

with the generalized coordinates $\mathbf{q}_\theta$ defined by the nodal values

$$\mathbf{q}_\theta = \begin{bmatrix}\theta|_1 & \theta|_2\end{bmatrix}^T\,. \tag{103}$$

To prevent the element from locking, a reduced numerical integration order of 5 (e.g., via 3 point Gauß' integration) is recommended when integrating bending and torsional stiffness terms over the beam's axis in Eq. (37), whereas the axial stiffness term shall be integrated exact, which means a numerical integration order of 9 or higher.

3.6 Director Update

For small deformation problems it is sufficient to consider a director $\mathbf{d}$, which is constant in space and time. However, problems arise, if the beam's axis, i.e., the axial slope $\mathbf{r}'(\xi)$ becomes (numerically) collinear with the director for any ξ. In this case the projection Eq. (99) becomes singular and the orientation of the beam's cross-section remains unknown. Note that the same holds not only for the deformed configuration, but also for the undeformed configuration. As a remedy, the director is chosen to vary

1. in space, achieved, e.g., by a spatial interpolation of $\mathbf{d}(\xi)$ between the two neighboring nodal directors $\mathbf{d}^1$ and $\mathbf{d}^2$, e.g., by a linear interpolation

$$\mathbf{d}(\xi) = S_5\mathbf{d}^1 + S_6\mathbf{d}^2\,,$$

2. in time, by performing an update of the nodal directors $\mathbf{d}^1$ and $\mathbf{d}^2$ past each time or load step, given as a function of the current orientation of the local frame at the ith node, e.g.,

$$\mathbf{d}^1(t_j) = \mathbf{e}_{30}(\xi = -\frac{L}{2}, t = t_{j-1}),$$
$$\mathbf{d}^2(t_j) = \mathbf{e}_{30}(\xi = +\frac{L}{2}, t = t_{j-1}),$$

or optionally by the post-rotated update

$$\mathbf{d}^1(t_j) = \mathbf{e}_3(\xi = -\frac{L}{2}, t = t_{j-1}), \qquad \theta^1(t_j) = 0$$
$$\mathbf{d}^2(t_j) = \mathbf{e}_3(\xi = +\frac{L}{2}, t = t_{j-1}), \qquad \theta^2(t_j) = 0$$

for all time steps t_j.

Let it be finally mentioned that the proposed Bernoulli–Euler beam finite element provides C^1-continuity along element borders only for the geometry of the beam axis, whereas the torsion of the cross-section, i.e., angle θ, is just C^0-continuous. Hence, the element has fourth-order convergence in problems with insignificant torsional effects, and a second-order convergence in all remaining problems. A fully C^1 continuous setting, requiring the rate of the torsional angle $\dot{\theta}$ to remain zero at the FE-nodes (in order to serve as a generalized coordinate) together with a conforming interpolation of the torsional angle θ (and the director $\mathbf{d}$) along the beam axis, is left for further investigation.

4 ANC Finite Elements with Shear and Cross-Section Deformation

In this section, the 2D and 3D formulations of thick ANC beam finite elements which include shear and cross-section deformation are discussed. In addition to the previous sections, displacements and displacement gradients are utilized rather than position and position gradients.

4.1 Kinematics of Thick Gradient-Deficient ANC Beam Finite Elements

Omar and Shabana (2001) presented an ANC finite element, in which a slope vector is used for modeling the shear deformation. For the 2D gradient-deficient ANC finite element, the latter finite element is modified by omitting the axial slope vector. The element is parametrized by displacements and displacement gradients at the nodes which form the degrees of freedom. Figure 2c shows a sketch of the fully parametrized element. The gradient-deficient element is obtained, if the axial slope vector $\mathbf{x}_{,\xi}$ is eliminated. The interpolation for a two-noded resp. a three-noded beam element is given with linear resp. quadratic shape functions. In case of the two-noded element, the shape functions are chosen according to Matikainen et al. (2009),

$$\begin{aligned} S_1 &= \frac{1}{L}\left(\frac{L}{2} - \xi\right), \quad S_2 = \eta S_1, \\ S_3 &= \frac{1}{L}\left(\frac{L}{2} + \xi\right), \quad S_4 = \eta S_3. \end{aligned} \tag{104}$$

In case of the three-noded element, the shape functions are chosen similar to those given by Mikkola et al. (2007) as

$$\begin{aligned} S_1 &= -\frac{2}{L^2}\xi\left(\frac{L}{2} - \xi\right), && S_2 = \eta S_1, \\ S_3 &= +\frac{2}{L^2}\xi\left(\frac{L}{2} + \xi\right), && S_4 = \eta S_3, \\ S_5 &= -\frac{4}{L^2}\left(\xi - \frac{L}{2}\right)\left(\xi + \frac{L}{2}\right), && S_6 = \eta S_5. \end{aligned} \tag{105}$$

The 3D gradient deficient ANC beam elements can be defined as the generalization of the 2D elements discussed above. Here, the two transverse slope vectors, which are in the cross-section plane, are used as degrees of freedom, compare Fig. 2d. In the spatial case, the shape functions of the linear (two-noded) element are given by

$$\begin{aligned} S_1(\xi,\eta,\zeta) &= \frac{1}{2} - \frac{\xi}{L}, \quad S_2(\xi,\eta,\zeta) = \eta S_1, \quad S_3(\xi,\eta,\zeta) = \zeta S_1, \\ S_4(\xi,\eta,\zeta) &= \frac{1}{2} + \frac{\xi}{L}, \quad S_5(\xi,\eta,\zeta) = \eta S_4, \quad S_6(\xi,\eta,\zeta) = \zeta S_4. \end{aligned} \tag{106}$$

The shape functions for the quadratic (three-noded) ANC beam finite element are given by

$$S_1 = -\frac{2}{L^2}\xi\left(\frac{L}{2} - \xi\right), \qquad S_2 = \eta S_1, \qquad S_3 = \zeta S_1,$$

$$S_4 = +\frac{2}{L^2}\xi\left(\frac{L}{2}+\xi\right), \qquad S_5 = \eta S_4, \qquad S_6 = \zeta S_4,$$
$$S_7 = -\frac{4}{L^2}\left(\xi-\frac{L}{2}\right)\left(\xi+\frac{L}{2}\right), \qquad S_8 = \eta S_7, \qquad S_9 = \zeta S_7. \tag{107}$$

4.2 Virtual Work of Elastic Forces for Thick Beams with Shear and Cross-Section Deformation

In addition to the structural mechanics formulation, which is customary for thin beams, a structural as well as a continuum mechanics based formulation is provided for shear and cross-section deformable ANC finite elements. Following the work of Gerstmayr et al. (2008), the work of elastic forces can be based on Reissner's nonlinear rod theory, see Reissner (1972), as implemented by Simo and Vu-Quoc (1986a), and a continuum mechanics based formulation, using a St. Venant–Kirchhoff material. For the 3D case, see Nachbagauer et al. (2011).

4.2.1 Continuum Mechanics Formulation

In the original shear-deformable ANC beam finite element by Omar and Shabana (2001), the elastic strain energy is defined using the Green's strain and the second Piola–Kirchhoff stress, as provided in Eq. (28). The main problem of this original continuum mechanics based formulation is the Poisson-locking phenomenon. In the original approach, the strain energy of a beam element with a rectangular cross-section is written in terms of the engineering strain vector $\bar{\varepsilon}$ and the elasticity matrix $\bar{\mathbf{D}}_{\mathrm{CM}}$ as presented in Eq. (33). The main problem of the original continuum mechanics based formulation arises since the Poisson ratio ν couples axial strains $\bar{E}_{11}$ and transverse normal strains $\bar{E}_{22}$ in the stress-strain relation. For pure axial deformation, the Poisson effect is modeled exactly. However, for bending deformation, the Poisson effect would require a trapezoidal deformation of the cross-section, which is not available in the original formulation. To avoid the locking effect, the strain energy is modified based on the idea of Gerstmayr et al. (2008). The elasticity matrix is split into two parts:

$$\bar{\mathbf{D}}_{\mathrm{CM}} = \bar{\mathbf{D}}^0_{\mathrm{CM}} + \bar{\mathbf{D}}^{\nu}_{\mathrm{CM}}, \tag{108}$$

in which $\bar{\mathbf{D}}^0_{\mathrm{CM}}$ does not include the Poisson ratio ν, while $\mathbf{D}^{\nu}$ involves the Poisson effect only. Hereafter, the strain energy is integrated over the cross-section, see Eq. (28), in which the part related to $\bar{\mathbf{D}}^0_{\mathrm{CM}}$ is integrated over the cross-section and the other part related to $\bar{\mathbf{D}}^{\nu}_{\mathrm{CM}}$ is integrated along the beam axis only using the cross-sectional area.

4.2.2 Structural Mechanics Formulation

The idea of the structural mechanics formulation is to incorporate the strain energy of classical nonlinear rod theories into the ANCF, for details see Gerstmayr et al. (2008) and Nachbagauer et al. (2011). The planar case of Eq. (37) reads

$$\delta W^{\text{int}} = \int_L EA\bar{\Gamma}_1 \delta\bar{\Gamma}_1 + k_s GA\bar{\Gamma}_2 \delta\bar{\Gamma}_2 + EI\bar{\kappa}\delta\bar{\kappa}\, d\xi, \tag{109}$$

in which the axial stiffness EA, the shear stiffness GA with the shear correction factor k_s, and the bending stiffness EI are coupled to the generalized strain measures for axial, shear, and bending strains, respectively. As proposed by Simo and Vu-Quoc (1986a), shear locking is eliminated by means of reduced integration here.

An additional term in the strain energy is necessary regarding the degrees of freedom of the cross-section deformation. Following Gerstmayr et al. (2008), the additional thickness strain energy $W_{\text{cs}}^{\text{int}}$ in case of a 2D beam finite element can be defined—in case of a rectangular cross-section—by

$$\delta W_{\text{cs}}^{\text{int}} = \int_L EA\bar{E}_{22}\, \delta\bar{E}_{22}\, d\xi, \tag{110}$$

which is defined similar to Eq. (48). The enhanced strain energy in the structural mechanics based formulation is the sum of the conventional strain energy W^{int} in Eq. (109) and $W_{\text{cs}}^{\text{int}}$ in Eq. (110), see Eq. (37). In the 3D case, the structural mechanics based formulation follows Simo (1985). For the case of simple symmetric cross-sections, see Eqs. (42) and (37) can be given for the single components,

$$\begin{aligned}\delta W^{\text{int}} = \int_L \; & EA\bar{\Gamma}_1 \delta\bar{\Gamma}_1 + GAk_2\bar{\Gamma}_2 \delta\bar{\Gamma}_2 + GAk_3\bar{\Gamma}_3 \delta\bar{\Gamma}_3 \\ & + GJk_t\bar{\kappa}_1 \delta\bar{\kappa}_1 + EI_2\bar{\kappa}_2 \delta\bar{\kappa}_2 + EI_3\bar{\kappa}_3 \delta\bar{\kappa}_3\, d\xi.\end{aligned} \tag{111}$$

In the 3D case, the virtual work of elastic forces covering cross-section deformation follows from Eq. (48).

5 Evaluation of the Accuracy and Performance of ANC Finite Elements

This section is dedicated to outline the numerical behavior of four of the proposed ANC beam finite elements, all of which are implemented in the open-source flexible multibody system dynamics code HOTINT,[2] see Gerstmayr et al. (2013a). Henceforth, let us use abbreviations as in Table 1.

[2] http://www.hotint.org/.

The interested reader is referred to the works by Gerstmayr and Irschik (2008), Gerstmayr et al. (2008), Nachbagauer et al. (2011), Nachbagauer et al. (2013) and Gruber et al. (2013), in which each of the proposed ANC beam finite elements is tested separately.

5.1 Static Example (Planar): Largely Deforming Cantilever

In this example we aim to compare the convergence and performance properties of all proposed ANC beam finite elements (Table 1) at once, i.e., both thin and thick elements are studied on behalf of the same setup.

A cantilever with length L and square cross-section with side-length a is subjected to a point load $\mathbf{F}$ acting at the material point B (which is the tip of the beam axis, see Fig. 4). The material parameters of the cantilever are defined by Young's modulus E and Poisson's ratio ν as

$$E = 2.07 \times 10^{11}\ \mathrm{N/m^2}\,, \qquad \nu = 0.3\,,$$

based on which the shear modulus G and the shear correction factor k_s are given by

$$G = \frac{E}{2(\nu+1)}\ \mathrm{N/m^2}\,, \qquad k_s = \frac{10(1+\nu)}{12+11\,\nu}\,. \tag{112}$$

Table 1 Types of ANC beam finite elements tested in Sect. 5

Name	Theory	Description
BE2D	Sect. 3	Thin beam in 2D (acc. to Bernoulli–Euler theory)
BE3D	Sect. 3.5	Same in 3D
SQ2D	Sect. 4	Shear deformable beam in 2D
SQ3D	Sect. 4	Same in 3D

Throughout the whole section the shear-deformable ANC beam finite elements SQ2D and SQ3D are considered to use quadratic shape functions, as defined in Eq. (105)

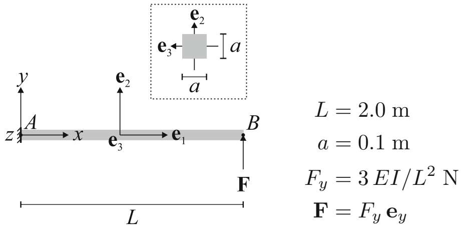


Fig. 4 Geometrical setup of the cantilever of Sect. 5.1 in reference configuration

A reference solution to the problem has been computed in the mathematical software framework Maple by solving an elliptic integral equation utilizing global polynomial shape functions, see Gerstmayr and Irschik (2008). The respective polynomial degree was chosen such that the first 12 digits of the displacement at material point B, reading

$$\mathbf{u}^{\mathrm{B}}_{\mathrm{ref}} = \begin{cases} -0.50853730436\,\mathbf{e}_x + 1.20723985455\,\mathbf{e}_y\,, & \text{for BE2D and BE3D}\,, \\ -0.50946471774\,\mathbf{e}_x + 1.20882282955\,\mathbf{e}_y\,, & \text{for SQ2D and SQ3D}\,, \end{cases} \tag{113}$$

have been converged.

All of the four beams, cf. Table 1, were tested in a scenario with ten uniform load steps, i.e.,

$$\mathbf{F}_i = \frac{i}{10}\mathbf{F}\,.$$

At each of those load steps a nonlinear system is solved by means of Newton's method, utilizing the solution of the previous load step as initial guess at the current load step. Newton's method is terminated if the relative error (i.e., max-norm of the actual residual over max-norm of the initial residual) becomes less than the bound $\varepsilon = 10^{-8}$. The overall performance and convergence behavior of the respective ANC beam finite elements are documented in Table 2 as well as in the convergence plots of Fig. 5 and a performance plot in Fig. 6.

Studying these tables and figures we arrive at the following conclusions:

1. All of the elements of Table 1 require roughly the same number of Newton iterations, independently of the underlying spatial refinement level.
2. Comparing the error of tip deflection $|\mathbf{u}^{\mathrm{B}}_{\mathrm{ref}} - \mathbf{u}^{\mathrm{B}}_{\mathrm{FE}}|$ versus number of elements (see also the left plot in Fig. 2), a quantitatively slightly different, but asymptotically equal behavior of all element types can be observed, namely a convergence order of 4 (meaning a decrease of the error roughly by a factor of c^{-4} if the number of elements is increased by a factor of $c > 0$.
3. The right plot in Fig. 2 seems to be a consequence from the left plot, owing to the fact that thick (i.e., shear-deformable) beam elements naturally own more degrees of freedom than their thin counterparts. The same holds of course with respect to the dimensionality of the several beam types.
4. The final plot in Fig. 6 shows that thin and thick elements need roughly the same computational time asymptotically, both in the planar and in the spatial case.

5.2 *Free Beam Flying in Plane*

By this planar dynamic benchmark example we aim to compare the computational speed of all the ANC finite elements presented in Table 1, as well as their convergence in terms of a displacement error, integrated over time.

Table 2 Performance table for the example *Largely deforming cantilever* of Sect. 5.1

Spatial discretization			Performance	
#NEL	#DOF	Err. (m)	CPU (s)	#Its.
SQ2D				
1	12	1.098e-001	0.109	51
2	20	1.516e-002	0.187	49
4	36	1.327e-003	0.297	49
8	68	9.159e-005	0.531	49
16	132	5.874e-006	0.952	49
32	260	3.695e-007	1.731	49
64	516	2.312e-008	3.244	50
BE2D				
1	8	2.585e-001	0.094	49
2	12	3.755e-002	0.156	49
4	20	2.035e-003	0.265	49
8	36	5.060e-005	0.655	49
16	68	1.202e-006	0.827	49
32	132	3.876e-008	1.716	49
64	260	1.820e-009	3.136	49
SQ3D				
1	27	1.098e-001	1.482	54
2	45	1.516e-002	2.324	55
4	81	1.327e-003	4.586	55
8	153	9.159e-005	8.471	55
16	297	5.874e-006	16.41	55
32	585	3.695e-007	32.6	55
64	1161	2.312e-008	70.86	55
BE3D				
1	14	2.585e-001	0.905	49
2	21	3.755e-002	1.404	49
4	35	2.035e-003	2.652	49
8	63	5.060e-005	4.695	49
16	119	1.202e-006	9.002	49
32	231	3.876e-008	17.94	49
64	455	1.820e-009	35.47	49

CPU-time in seconds (CPU (s)) and number of Newton iterations (#Its.) for various levels of spatial approximation including number of elements (#NEL), total degrees of freedom (#DOF), and approximation error (Err. (m)), measured by the error of the tip deflection, i.e. Err. $= |\mathbf{u}^{\mathrm{B}}_{\mathrm{ref}} - \mathbf{u}^{\mathrm{B}}_{\mathrm{FE}}|$

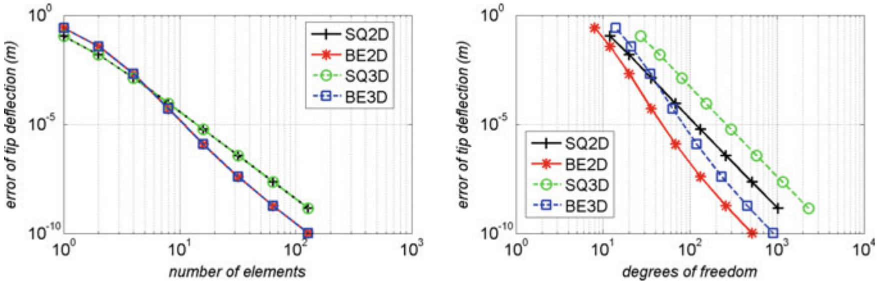


Fig. 5 Convergence plot of the example problem in Sect. 5.1 showing the error of tip deflection $|\mathbf{u}^{\mathrm{B}}_{\mathrm{ref}} - \mathbf{u}^{\mathrm{B}}_{\mathrm{FE}}|$ versus number of elements (*left*) and degrees of freedom (*right*)

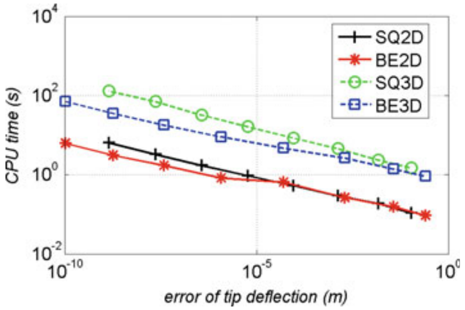
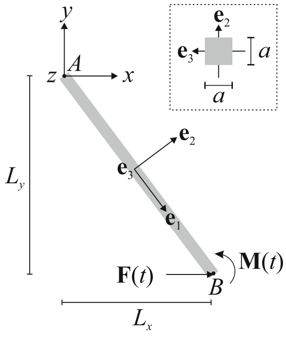


Fig. 6 The performance of the ANC beam finite elements in the example problem of Sect. 5.1 is compared in terms of CPU-time versus error of tip deflection $|\mathbf{u}^{\mathrm{B}}_{\mathrm{ref}} - \mathbf{u}^{\mathrm{B}}_{\mathrm{FE}}|$



$$L_x = 0.6\ \mathrm{m}$$
$$L_y = 0.8\ \mathrm{m}$$
$$a = 0.05\ \mathrm{m}$$

$$f(t) = \begin{cases} 2\,t & t \in [0, 0.5[\,, \\ 1 & t \in [0.5, 2.5[\,, \\ 2\,(3-t) & t \in [2.5, 3[\,, \\ 0 & t \geq 3 \end{cases}$$

$$F_x = 0.3\ \mathrm{N}\,, \qquad \mathbf{F}(t) = f(t)\, F_x\, \mathbf{e}_x$$
$$M_z = 0.3\ \mathrm{Nm}\,, \quad \mathbf{M}(t) = f(t)\, M_z\, \mathbf{e}_z$$

Fig. 7 Geometrical setup of the free beam of Sect. 5.2 in reference configuration

A free beam with a square cross-section, as shown in Fig. 7, is subjected to a force $\mathbf{F}(t)$ and a moment $\mathbf{M}(t)$, both acting over the time $t \in [0, 10]$ at the beam axis point B. The material parameters of the beam are defined by Young's modulus E, Poisson's ratio ν, and the material density ρ as

$$E = 1 \times 10^5 \text{ N/m}^2 , \qquad \nu = 0.3 , \qquad \rho = 2500 \text{ kg/m}^3 , \tag{114}$$

based on which the shear modulus G and the shear correction factor k_s are computed up to double precision, utilizing Eq. (112).

Note that the problem setup is defined similar but not equal to the well-known *flying spaghetti problem* of Simo and Vu-Quoc (1986b). The reason for considering not the original but a modified problem setup is to end up with less differences in the kinematic behavior of thin and shear-deformable beam elements.

Time integration is numerically performed by a uniform time step size $\Delta t = 0.001$ s utilizing a two staged trapezoidal rule (Lobatto), which means three integration points per time step and a fourth-order global convergence in time. In difference to the test example *largely deforming cantilever* of Sect. 5.1, where a classical Newton method was used per load step, we choose to update the Jacobian not at each time step, but only if required, i.e., if the Newton residual is not converging fast enough. Although generally resulting in much more iteration steps, such a modified Newton method speeds up the simulation, particularly if the considered time steps are comparatively small. Alike the classical Newton method, the modified Newton method is terminated, if the relative error (i.e., max-norm of the actual residual over max-norm of the initial residual) becomes less than the bound $\varepsilon = 10^{-8}$.

Two convergence plots in Fig. 8 and a performance plot in Fig. 9 conclude this example. In all of these plots, the integrated deflection error

$$\epsilon_T = \left(\int_0^T |\mathbf{u}^{\mathrm{B}}_{\mathrm{ref}} - \mathbf{u}^{\mathrm{B}}_{\mathrm{FE}}|^2 dt \right)^{1/2} \tag{115}$$

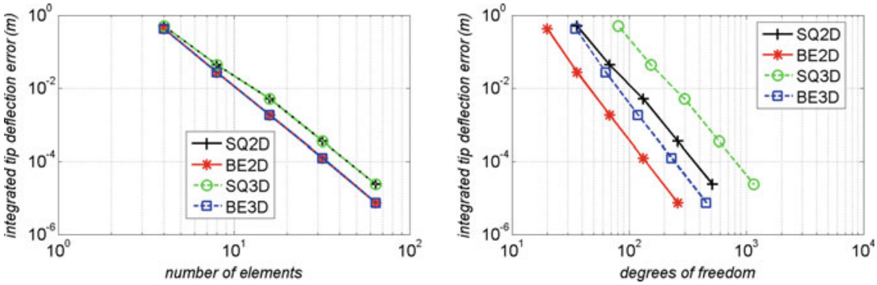


Fig. 8 Convergence plot of the example problem in Sect. 5.2 showing the integrated tip deflection error, as defined in Eq. (115), versus number of elements (*left*) or degrees of freedom (*right*)

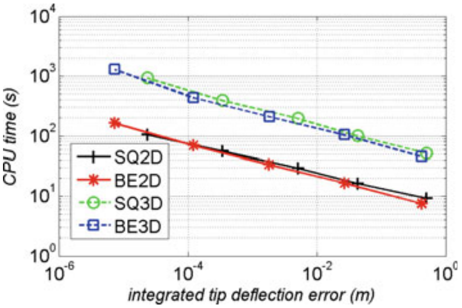


Fig. 9 The performance of the ANC beam finite elements in the example problem of Sect. 5.2 is compared in terms of CPU-time versus integrated tip deflection error, as defined in Eq. (115)

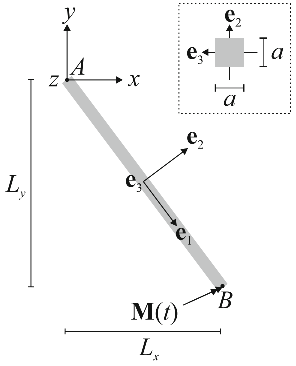
at material point B (see Fig. 7) served as a measure of the finite element approximation error. The reference solution $\mathbf{u}^{\mathrm{B}}_{\mathrm{ref}}$ (which is different for thin and for shear deformable beams) is computed by means of a highly refined FE-solution of 128 elements, also at a time step size of $\Delta t = 0.001$ s.

5.3 *Free Beam Flying in Space*

If a spatial beam formulation uses angular in addition to absolute coordinates (which is the case in the beam class BE3D, but not so in the beam class SQ3D), the shape function interpolation of those angular coordinates (evaluated at the integration points along the beam's axis) causes the total energy of the beam to be no more conserved, and thus any time integration scheme becomes unstable and must fail to converge. To demonstrate this effect in this third test example, we consider a similar problem setup as in Sect. 5.2, however with slightly different material and geometrical parameters, and with a different loading scenario causing the beam to move out of plane.

A free beam with a square cross-section, as shown in Fig. 10, is subjected to a moment $\mathbf{M}(t)$ acting at the time t at the beam axis point B. The material parameters of the beam are defined by Young's modulus E, Poisson's ratio ν, and the material density ρ as in Eq. (114), based on which the shear modulus G and the shear correction factor k_s are computed up to double precision, utilizing Eq. (112).

Although the beam class BE3D shows better convergence compared to the beam class SQ3D (as shown in Fig. 12) the simulation with type BE3D elements would become unstable after a while. To be precise, an implicit time integration scheme (of type Lobatto using 2 stages and a uniform time step size of 0.001 s) would fail to converge after 8 s of simulation time when using a spatial discretization with 8 elements, after 9 s when using 16 elements, and after 12 s when using 32 elements



$$L_x = 0.6\ \mathrm{m}$$
$$L_y = 0.8\ \mathrm{m}$$
$$a = 0.02\ \mathrm{m}$$
$$f(t) = \begin{cases} 2\,t & t \in [0, 0.5[\,, \\ 1 & t \in [0.5, 2.5[\,, \\ 2\,(3-t) & t \in [2.5, 3[\,, \\ 0 & t \geq 3 \end{cases}$$
$$M_x = \frac{1}{72}\ \mathrm{Nm}\,, \quad M_y = \frac{1}{100}\ \mathrm{Nm}$$
$$\mathbf{M}(t) = f(t)\,(M_x\,\mathbf{e}_x + M_y\,\mathbf{e}_y)$$

Fig. 10 Geometrical setup of the free beam of Sect. 5.3 in reference configuration

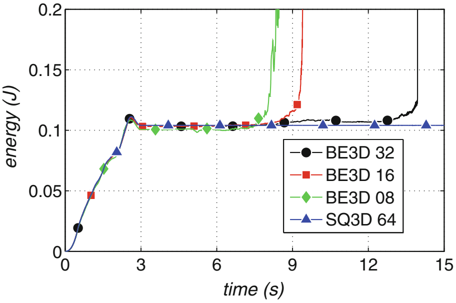


Fig. 11 Sum of kinetic and potential energy of the beam computed with a uniform time step of 0.001 s and a spatial discretization of 8, 16, and 32 elements of type BE3D compared to 64 elements of type SQ3D

of beam type BE3D, whereas simulations with the same type of integration scheme but using SQ3D elements did not show instability effects at all (at least, in our tests the simulation remained stable until 50 s of simulation time).

This issue becomes even more evident if we study the total (i.e., the sum of kinetic and potential) energy, see Fig. 11. Analytically the total energy of the flying beam must stay constant as soon as the outer forces, i.e., the tip moment, become zero (which happens past 3 s of simulation time, see the definition of the time ramp $f(t)$ in Fig. 10). In case of simulations with BE3D elements, sudden energy blowups occur whereas in simulations with SQ3D elements the total energy is conserved, independently of the spatial refinement.

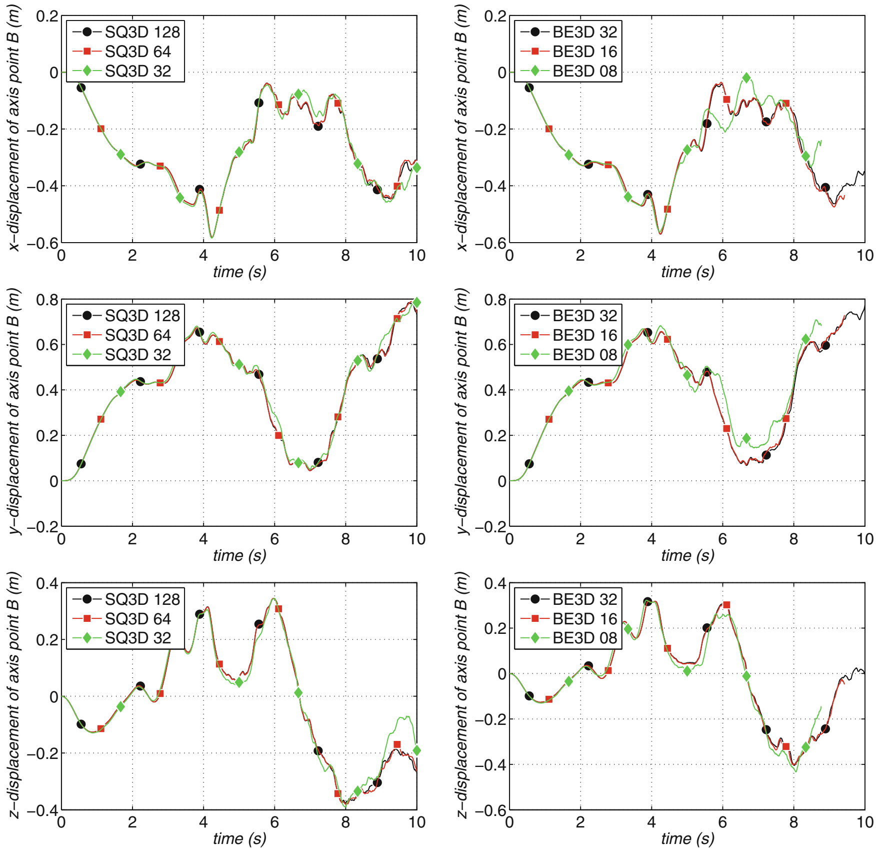


Fig. 12 Displacement in x-, y-, and z-direction of the axial point B of the beam in Sect. 5.3 computed with 32, 64, and 128 elements of type SQ3D (*left*), and 8, 16, and 32 elements of type BE3D (*right*)

6 Conclusions

In the present chapter, the absolute nodal coordinate formulation has been introduced and some specific finite elements, which are based upon this formulation, have been presented in a unified notation regarding kinematics and work of elastic forces. The finite elements under investigation have been studied regarding its convergence as well as the stability. It turned out that displacement-based finite elements, which do not employ rotations as degrees of freedom, are not showing numerical instabilities as compared to those which contain at least one rotational parameter. Finite elements with rotational parameters, however, have other advantages. For those elements, it is necessary to obtain stable numerical integration schemes, which are discussed in detail in other chapters of this book.

Acknowledgments The authors Gerstmayr, Humer, and Gruber have been supported by the Linz Center of Mechatronics (LCM) in the framework of the Austrian Comet-K2 programme. K. Nachbagauer acknowledges support from the Austrian Science Fund (FWF): T733-N30.

References

Antman, S. S. (1972). The theory of rods. In S. Flügge & C. Truesdell (Eds.), *Handbuch der Physik* (Vol. VIa/2, pp. 641–703). Berlin: Springer.

Berzeri, M., & Shabana, A. A. (2002). Development of simple models for the elastic forces in the absolute nodal co-ordinate formulation. *Journal of Sound and Vibration*, *235*(4), 539–565.

Betsch, P., & Steinmann, P. (2003). Constrained dynamics of geometrically exact beams. *Computational Mechanics*, *31*, 49–59.

Dibold, M., Gerstmayr, J., & Irschik, H. (2009). A detailed comparison of the absolute nodal coordinate and the floating frame of reference formulation in deformable multibody systems. *ASME Journal of Computational and Nonlinear Dynamics*, *4*(2), 10.

Dmitrochenko, O. N., & Pogorelov, D. Y. (2003). Generalization of plate finite elements for absolute nodal coordinate formulation. *Multibody System Dynamics*, 10, 17–43. http://dx.doi.org/10.1023/A:1024553708730. ISSN: 1384-5640.

Escalona, J. L., Hussien, H. A., & Shabana, A. A. (1998). Application of the absolute nodal coordinate formulation to multibody system dynamics. *Journal of Sound and Vibration*, *214*(5), 833–851.

Frischkorn, J., & Reese, S. (2012). A novel solid-beam finite element for the simulation of nitinol stents. In *Proceedings of the ECCOMAS 2012 European Congress on Computational Methods on Applied Sciences and Engineering, Vienna, Austria*.

Gerstmayr, J. (2009). A corotational approach for 3D absolute nodal coordinate elements. In *Proceedings of the ASME IDETC/CIE 2009, San Diego, USA*.

Gerstmayr, J., & Irschik, H. (2003). Vibrations of the elasto-plastic pendulum. *International Journal of Nonlinear Mechanics*, *38*, 111–122.

Gerstmayr, J., & Irschik, H. (2008). On the correct representation of bending and axial deformation in the absolute nodal coordinate formulation with an elastic line approach. *Journal of Sound and Vibration*, *318*, 461–487.

Gerstmayr, J., & Matikainen, M. K. (2006). Analysis of stress and strain in the absolute nodal coordinate formulation. *Mechanics Based Design of Structures and Machines*, *34*, 409–430.

Gerstmayr, J., & Shabana, A. A. (2006). Analysis of thin beams and cables using the absolute nodal coordinate formulation. *Nonlinear Dynamics*, *45*(1–2), 109–130.

Gerstmayr, J., Matikainen, M. K., & Mikkola, A. M. (2008). A geometrically exact beam element based on the absolute nodal coordinate formulation. *Journal of Multibody System Dynamics*, *20*, 359–384.

Gerstmayr, J., Dorninger, A., Eder, R., Gruber, P., Reischl, D., Saxinger, M., Schörgenhumer, M., Humer, A., Nachbagauer, K., Pechstein, A., & Vetyukov, Y. (2013a). Hotint: A script language based framework for the simulation of multibody dynamics systems. In *9th International Conference on Multibody Systems, Nonlinear Dynamics, and Control* (Vol. 7B). doi:10.1115/DETC2013-12299.

Gerstmayr, J., Sugiyama, H., & Mikkola, A. (2013b). Review on the absolute nodal coordinate formulation for large deformation analysis of multibody systems. *ASME Journal of Computational and Nonlinear Dynamics*, *8*, 031016 (12 pages).

Gruber, P. G., Nachbagauer, K., Vetyukov, Y., & Gerstmayr, J. (2013). A novel director-based Bernoulli-Euler beam finite element in absolute nodal coordinate formulation free of geometric singularities. *Mechanical Sciences*, *4*(2), 279–289. doi:10.5194/ms-4-279-2013. http://www.mech-sci.net/4/279/2013/.

Irschik, H., & Gerstmayr, J. (2009a). A hyperelastic Reissner-type model for non-linear shear deformable beams. In I. Troch & F. Breitenecker (Eds.), *Proceedings of the Mathmod 09*.

Irschik, H., & Gerstmayr, J. (2009b). A continuum mechanics based derivation of Reissner's large-displacement finite-strain beam theory: The case of plane deformations of originally straight Bernoulli-Euler beams. *Acta Mechanica*, *206*, 1–21.

Irschik, H., & Gerstmayr, J. (2011). A continuum-mechanics interpretation of Reissner's non-linear shear-deformable beam theory. *Mathematical and Computer Modelling of Dynamical Systems*, *17*(1), 19–29.

Lan, P., & Shabana, A. (2010a). Integration of B-spline geometry and ANCF finite element analysis. *Nonlinear Dynamics*, *61*(1–2), 193–206.

Lan, P., & Shabana, A. (2010b). Rational finite elements and flexible body dynamics. *ASME Journal of Vibration and Acoustics*, *132*(4), 041007.

Matikainen, M. K., von Hertzen, R., Mikkola, A., & Gerstmayr, J. (2009). Elimination of high frequencies in the absolute nodal coordinate formulation. In *Proceedings of the Institution of Mechanical Engineers, Part K, Journal of Multi-body Dynamics*.

Mikkola, A. M., & Shabana, A. A. (2003). A non-incremental finite element procedure for the analysis of large deformations of plates and shells in mechanical system applications. *Multibody System Dynamics*, *9*, 283–309.

Mikkola, A. M., Garcia-Vallejo, D., & Escalona, J. L. (2007). A new locking-free shear deformable finite element based on absolute nodal coordinates. *Nonlinear Dynamics*, *50*, 249–264.

Nachbagauer, K., Pechstein, A. S., Irschik, H., & Gerstmayr, J. (2011). A new locking-free formulation for planar, shear deformable, linear and quadratic beam finite elements based on the absolute nodal coordinate formulation. *Multibody System Dynamics*, *26*, 245–263.

Nachbagauer, K., Gruber, P., & Gerstmayr, J. (2013). Structural and continuum mechanics approaches for a 3D shear deformable ANCF beam finite element: Application to static and linearized dynamic examples. *Journal of Computational and Nonlinear Dynamics*, *8*, 021004.

Olshevskiy, A., Dmitrochenko, O., & Kim, C.-W. (2013). Three-dimensional solid brick element using slopes in the absolute nodal coordinate formulation. *Journal of Computational and Nonlinear Dynamics*, *9*(2), 021001.

Omar, M. A., & Shabana, A. A. (2001). A two-dimensional shear deformable beam for large rotation and deformation problems. *Journal of Sound and Vibration*, *243*(3), 565–576.

Pechstein, A., & Gerstmayr, J. (2013). A Lagrange-Eulerian formulation of an axially moving beam based on the absolute nodal coordinate formulation. *Multibody System Dynamics*, *30*(3), 343–358.

Reissner, E. (1972). On one-dimensional finite-strain beam theory: The plane problem. *Journal of Applied Mathematics and Physics*, *23*, 795–804.

Reissner, E. (1973). On one-dimensional large-displacement finite-strain beam theory. *Studies in Applied Mathematics, LI*, *I*(2), 87–95.

Shabana, A. A., & Schwertassek, R. (1997). Equivalance of the floating frame of reference approach and finite element formulations. *International Journal of Non-Linear Mechanics*, *33*(3), 417–432.

Simo, J. C. (1985). A finite strain beam formulation. The three-dimensional dynamic problem. Part I. *Computer Methods in Applied Mechanics and Engineering*, *49*, 55–70.

Simo, J. C., & Vu-Quoc, L. (1986a). On the dynamics of flexible beams under large overall motions–the plane case: Part I. *Journal of Applied Mechanics*, *53*(4), 849–854. doi:10.1115/1.3171870. http://dx.doi.org/10.1115/1.3171870. ISSN: 0021-8936.

Simo, J. C., & Vu-Quoc, L. (1986b). On the dynamics of flexible beams under large overall motions–the plane case: Part II. *Journal of Applied Mechanics*, *53*(4), 855–863. doi:10.1115/1.3171871. http://dx.doi.org/10.1115/1.3171871. ISSN: 0021-8936.

Simo, J. C., & Vu-Quoc, L. (1986c). A three-dimensional finite-strain rod model. Part II: Computational aspects. *Computer Methods in Applied Mechanics and Engineering*, *58*, 79–116.

Simo, J. C., & Vu-Quoc, L. (1988). On the dynamics in space of rods undergoing large motions-a geometrically exact approach. *Computer Methods in Applied Mechanics and Engineering*, *66*, 125–161.

Sugiyama, H., & Shabana, A. A. (2004). Application of plasticity theory and absolute nodal coordinate formulation to flexible multibody system dynamics. *ASME Journal of Mechanical Design*, *126*, 478–487.

Sugiyama, H., Escalona, J. L., & Shabana, A. A. (2003). Formulation of three-dimensional joint constraints using the absolute nodal coordinates. *Nonlinear Dynamics*, *31*, 167–195.

Yakoub, R. Y., & Shabana, A. A. (2001). Three dimensional absolute nodal coordinate formulation for beam elements. *ASME Journal of Mechanical Design*, *123*, 606–621.

A Brief Introduction to Variational Integrators

Adrián J. Lew and Pablo Mata A

Abstract In this chapter, a brief introduction to the formulation of variational methods for finite-dimensional Lagrangian systems is presented. To this end, the first two sections focus on describing the Lagrangian and Hamiltonian points of view of mechanics for systems evolving on manifolds. Special attention is paid to the construction of the Lagrangian function and to the role of Hamilton's variational principle in the deduction of the balance equations. The relation between the symmetries of the Lagrangian function and the existence of invariants of the dynamics along with the symplectic nature of the flow are also addressed. In the third section, the discussion turns towards the formulation of a time-discrete analogue of the theory. The cornerstone of such a construction is given by a discrete analogue of Hamilton's variational principle which provides a systematic procedure to construct discrete approximations to the exact trajectory of a mechanical system on both the configuration space and the phase space. The approximation properties and the geometric characteristics of the resulting discrete trajectories are explained. Finally, we apply the variational methodology to construct symplectic and momentum-conserving time integrators for two problems of practical interest in engineering and science.

A.J. Lew (✉)
Department of Mechanical Engineering, Stanford University, Stanford,
CA 94305-4040, USA
e-mail: lewa@stanford.edu

P. Mata A
Centro de Investigación En Ecosistemas de la Patagonia,
CIEP Universidad Austral de Chile, Km 4.5 Camino Aysén, Coyhaique, Chile
e-mail: pmata@ciep.cl

P. Betsch (ed.), *Structure-preserving Integrators in Nonlinear Structural Dynamics and Flexible Multibody Dynamics*, CISM International Centre
for Mechanical Sciences 565, DOI 10.1007/978-3-319-31879-0_5

1 Introduction

1.1 Overview

A bumper sticker explaining how to construct variational integrators (VI) would read

> "Approximate the action instead of the equations of motion."

This simple idea turns out to be very powerful, as we shall have the opportunity to explore in these notes. In fact, it has underpinned the solutions to elastostatics problems with the finite element method for 60 years now. Why, when, and how an approximation of a fundamental object in classical mechanics, the action, gives rise to a convergent scheme to integrate the equations of motion are questions that we shall address.

To explain the implications of the above bumper sticker, in the following sections we briefly review the Lagrangian formulation of the mechanics of a conservative system, and then we mimic this process at the discrete level to construct variational integrators.

1.2 Perspective

Early approaches toward the creation of time integrators for ordinary differential equations (ODE's) consisted in constructing a suitable discretization of equations, without accounting for additional structure these equations might have. See for example, (Hughes 1987; Bathe 1996; Hairer et al. 1993; Hairer and Wanner 1996; Holmes 2007) among others.

An alternative point of view is given by the formulation of the so-called *structure-preserving* methods. Those methods are designed to preserve the geometric properties of the flow of the differential equations. This category includes but is not limited to:

- Methods that conserve the invariants of the dynamics such as energy-conserving integrators applied to conservative systems (Bayliss and Issacson 1975; Labudde and Greenspan 1976) or energy-momentum conserving methods (Simo et al. 1992; Simo and Wong 1991) which conserve energy as well as linear and angular momentum. These ideas have been also applied to problems described by partial differential equations (PDE's), such as for example to the dynamics of nonlinear solids (Gonzalez 2000). The list of contributions in this area is long, with for example (Armero and Romero 2001; Armero and Petoz 1999; Bauchau and Bottasso 1999; Betsch and Uhlar 2007; Borri et al. 2001), to name only a few of them.
- Numerical methods for dynamical systems evolving on general manifolds rather than on linear spaces. The key feature of these methods consists in that the resulting discrete trajectory belongs to the same configuration manifold as the time-continuous system (Iserles 1997; Desbrun et al. 2014). Many important problems

in physics are described in terms of dynamical systems evolving on *Lie-groups* (see Sect. A.2) such as for example, the dynamics in space of N-interacting rigid bodies, the dynamics of slender structures such as rods and filaments, the dynamics of shells and the motion of incompressible and inviscid fluids among others. Lie-group methods are numerical integrators specifically formulated for dynamical problems on Lie groups or on manifolds acted upon by Lie groups (Iserles et al. 2000; Celledoni et al. 2014).The application of these methods to the full-body problem can be consulted in (Lee et al. 2007; Celledoni and Owren 2003), to the dynamics of elastic and inelastic rods in (Mata 2015; Mata et al. 2008, 2009; Romero and Armero 2002, Simo et al. 1995) and to the dynamics of shells e.g., in (Simo and Tarnow 1994; Sansour and Wagner 2003). An error analysis of Lie-group methods can be consulted in (Faltinsen 2000).

- Symmetric methods for reversible problems. Reversible dynamical systems are characterized by the fact that inverting the direction of the velocity vector while keeping the initial position fixed, results in an inversion of the solution trajectory. Conservative mechanical systems are reversible. Numerical methods able to generate reversible numerical flows when applied to a reversible differential equations constitute an active field of research in geometric time integration (Hairer et al. 2006, Chap. V; Cano and Sanz-Serna 1988).
- Symplectic methods for Lagrangian/Hamiltonian problems. The Lagrangian/ Hamiltonian systems are among the most important dynamical systems in sciences and engineering. As noted in (Feng and Qin 2010) any conservative real physical process can be formulated as a Hamiltonian system, whether they have finite or infinite degrees of freedom. An outstanding property of Hamiltonian systems is the symplectic nature of their flows on the phase space. See Sects. 2.3 and 3.5 for a precise definition of continuous and discrete symplecticity. Examples of Hamiltonian systems appearing in science and engineering include but are not limited to the structural biology (Gay-Balmaz et al. 2009), molecular dynamics (Stavros 2014; Manning and Maddocks 1999), mathematical models in ecosystem dynamics (Kirwan 2008), superconductivity (Bogolyubov 1972), plasma physics (Larsson 1996), celestial mechanics and cosmology (Arnold et al. 2006), fluid mechanics (Desbrun et al. 2014; Gawlik et al. 2011), mechanics of materials and structures (Simo et al. 1988), theoretical physics (Esposito et al. 2004; Marsden 1988), aerospace engineering (Kasdin et al. 2005), satellite dynamics and control (Kuang et al. 2003; Koon et al. 2011), kinematics and dynamics of mechanisms and robots (Macchelli et al. 2009; Chen 1990) and other areas of seismic (Luo et al. 2013), mechanical and electrical (Clemente-Gallardo and Scherpen 2003) engineering. Symplectic integrators are methods specially formulated to produce a symplectic flows on the phase space. This property is intimately related to the ability of these methods to reproduce the long-time structure of the solutions of Hamiltonian ODE's (e.g., limit cycles, attractors, invariant manifolds, etc.) as it has been reported in several occasions e.g., (Bou-Rabee and Marsden 2009). A survey on symplectic time integration of Hamiltonian ODE's and Hamiltonian PDE's can be consulted in (Leimkuhler and Reich 2005; Feng and Qin 2010) and (Hairer et al. 2006, Chap. VI). The nonlinear stability of symplectic integrators

is considered in (McLachlan et al. 2004). The role of symplectic integration in optimal control is reviewed in (Chyba et al. 2009).

Therefore, formulating numerical methods able to preserve the geometric structure of the solutions of Hamiltonian systems will have extremely broad applications. This chapter focuses on describing a particular methodology, based on the discretization of a fundamental principle in mechanics, to construct structure-preserving methods for Lagrangian systems with symmetries.

1.3 Variational Integration

Variational integrators (VI's) constitute a more recent approach toward the creation of structure-preserving methods for finite-dimensional Lagrangian systems or for appropriate discretizations of some Hamiltonian continuum systems. Their construction is based on the formulation of a discrete analogue to Hamilton's variational principle. The basic idea consists in constructing a time-discrete approximation of the action integral called the discrete action sum. Stationary points of the discrete action sum are then the discrete-in-time trajectories of the mechanical system, and can be proved to approximate the exact trajectories as the time-step goes to zero. These ideas were originally explored in the works of (Maeda 1980; Maeda 1982) and of (Veselov 1988; Moser and Veselov 1991) in the context of integrable systems in mechanics.

The procedure used to construct a time-discrete trajectory defines a variational time integrator that shows a number of remarkable properties among which are:

(i) It conserves the invariants of the dynamics associated to the symmetries of the original mechanical system if the discrete action sum is designed to preserve the same symmetries. See Sects. 2.3 and 3.4. This is also known as a discrete version of Noether's theorem, see e.g., (Marsden and West 2001; Lew et al. 2004, 2003).

(ii) A discrete version of the Legendre transform allows to construct an alternative but otherwise equivalent form of the method that defines a discrete symplectic flow over the phase space. See Sect. 3.5.

(iii) The discrete trajectory displays an outstanding energy behavior. To be more precise, the value of the energy computed over the discrete trajectory remains close to its initial value for very long times, provided the time step is small enough, see e.g., (Leimkuhler and Reich 2005; Hairer et al. 2006).

(iv) Symplectic and momentum-conserving methods of arbitrarily high order of accuracy for dynamical systems evolving on nonlinear manifolds can be systematically constructed following the standard methodology of variational integration.

The essential aspects of VI's can be reviewed in (Wendlandt and Marsden 1997; Marsden and Wendlandt 1997; Marsden and West 2001) and in (Leok and Shingel 2012a; Leok 2005) general techniques for constructing variational integrators are provided. Spectral variational integrators are described in (Hall and Leok 2014a) and prolongation–collocation methods in (Leok and Shingel 2012b).

The variational methodology has been successfully applied to a broad range of fields such as for example:

- *Mechanical problems with multisymplectic geometry*. In (Marsden et al. 1998) a geometric-variational approach to continuous and discrete field theories is described and in (Marsden et al. 2001) the authors present a variational and multisymplectic formulation of both compressible and incompressible models of continuum mechanics. Asynchronous variational time integrators for finite element discretizations of deformable solids and field theories are formulated in (Lew et al. 2004, 2003; Lew 2003; Kale and Lew 2006; De León et al. 2008). An analysis of the stability properties of asynchronous VI's can be found in (Fong et al. 2008). See (Focardi and Maria-Mariano 2008) for a converge analysis of asynchronous VI's in linear elastodynamics and (Patrick and Cuell 2009) for a complete error analysis.

- *Dynamical systems evolving on nonlinear manifolds*. Discrete analogues of Euler–Poincaré and Lie–Poisson reduction theory for systems on finite-dimensional Lie groups with symmetries are developed in (Marsden et al. 1999). A Lie Poisson structure for a discrete mechanical system evolving on a Lie group is deduced in (Marsden et al. 2000). Lie group VI's applied to the full-body problem are formulated in (Lee et al. 2007) and VI's on two-spheres in (Lee et al. 2009). The extension of spectral variational integrators to Lie groups is developed in (Hall and Leok 2014b).

- *Structural elements: beams, rods, plates, and shells*. An explicit, second-order accurate VI that can be identified with a Lie-group, symplectic, partitioned Runge–Kutta method for finite element discretizations of geometrically exact rods is presented in (Mata 2015). The formulation of VI's for spatial beams and plates is carried out in (Demoures 2012; Demoures et al. 2014). In (Nichols and Murphey 2008) a VI for simulating the dynamics of cable structures is formulated. A discrete model for shells is formulated in (Grispun et al. 2003).

- *Contact and impact*. In (Fetecau et al. 2003b; Fetecau 2003) the classical theory of (smooth) Lagrangian mechanics is extended to the nonsmooth case in order to include collisions and the foundations of the multisymplectic formulation of nonsmooth continuum mechanics are presented in Fetecau et al. (2003a). An example of asynchronous collision integrators can be consulted in Wolff and Bucher (2013) and an application to polymer chains in (Leyendecker et al. 2012). In (Ryckman and Lew 2010, 2011, 2012) a new explicit dynamic contact algorithm that takes advantage of a variational asynchronous time integrator is formulated. A variational formulation of contact is formulated in (Harmon et al. 2009) and an optimization of this method is carried out in (Ainsley et al. 2012). See also (Cirak and West 2005).

- *Multibody dynamics and control*. A comparison between the numerical performance of VI's and energy-momentum schemes when applied to the numerical simulation of flexible multibody dynamics is presented in Betsch et al. (2010). The solvability of some geometric integrators for multibody systems is analyzed in (Kobilarov 2014). A discontinuous version of VI's is formulated in (Johnson et al. 2014) to treat collisions in multibody systems. In (Jiménez et al. 2013) numerical methods for optimal control of mechanical systems in the Lagrangian setting are formulated.

- *Stochastic differential equations*. In (Bou-Rabee and Owhadi 2009) a continuous and discrete Lagrangian theory for stochastic Hamiltonian systems on manifolds is presented. See also (Wang et al. 2009; Wang 2007). The long-time statistical properties of a Lie–Trotter splitting for inertial Langevin equations are presented in (Bou-Rabee and Owhadi 2010). The splitting is defined as a composition of a variational integrator with an Ornstein–Uhlenbeck flow (Van Bargena and Dimitroff 2009). Further material about geometric integrators for stochastic dynamical systems can be found in Tao et al. (2010). Variational integrators for constrained, stochastic mechanical systems are presented in Bou-Rabee and Owhadi (2007).

- *Constrained and forced problems*. The formulation of variational integrators applied to Lagrangian systems subjected to holonomic constraints or forces can be consulted in (Marsden and West 2001; West 2004). A variational discrete null space method is proposed in (Leyendecker et al. 2008). See also (Leyendecker et al. (2007)). A study of the Γ-convergence of VI's to the corresponding continuum action functional and the convergence properties of the discrete trajectories to stationary points of the continuum problem is presented in Schmidt et al. (2009). The use of a discrete version of Lagrange–D'Alembert principle allows to include non-conservative generalized forces. This is particularly useful for weakly dissipative systems (Kane et al. 1999, 2000).

- *Dynamics of fluids*. A geometric theory for fluid dynamics can be found in (Marsden and Ratiu 1999; Arnold and Khesin 1998; Pavlov 2009). Traditionally, numerical methods for fluid dynamics have been rarely designed to preserve the geometric structure of the solution trajectories, resulting in the introduction of spurious numerical artifacts. In contrast, in (Pavlov et al. 2011) discrete equations of motion for fluid dynamics are derived from first principles in Eulerian form. In (Gawlik et al. 2011) a variational discretization of continuum theories arising in fluid dynamics, magnetohydrodynamics and the dynamics of complex fluids is presented and in (Desbrun et al. 2014) a structure-preserving scheme for the dynamics of rotating and/or stratified fluids is formulated.

- *Thermoelasticity and nonequilibrium thermodynamics*. The dynamics of systems undergoing irreversible processes has motivated the formulation of structure-preserving integrators able to satisfy the first (energy conservation) and second (nondecreasing entropy of an isolated system) laws of thermodynamics along with the conservation of the invariants associated to the symmetries of the system. Those methods are frequently called *thermodynamically consistent*. See e.g., (Romero 2009; Bargmann and Steinmann 2008a, b). The simplest type of such systems are likely to be the thermomechanical systems. A Lagrangian/Hamiltonian formulation of thermoelasticity is obtained by introducing the concept of thermal displacements (Maugin 2000; Maugin and Kalpakides 2002) and (Green and Naghdi 1993, 1991, 1995). In this formulation, the temperatures are obtained as the time derivatives of the thermal displacements and the entropy is the conjugate momentum to the temperature. In (Mata and Lew 2011) a class of variational integrators for finite-dimensional adiabatic thermoelastic is formulated. The same ideas are applied in (Mata and Lew 2014) to develop thermodynamically consistent methods for finite element discretizations of deformable elastic solids with second sound. Unfortunately, it is not possible to construct a Hamiltonian formulation of thermoelasticity with heat conduction of Fourier type.[1] In (Mata and Lew 2012), an entropy flux term of Fourier type is added as a dissipative perturbation to the Hamiltonian form of the balance equations of adiabatic thermoelasticity and a discrete version of D'Alembert's principle is used to formulate structure-preserving methods. A similar approach has been followed in (Kern et al. 2014).

Other applications of variational integration can be reviewed, e.g., in (Kharevych et al. 2006; Kraus 2013; Ober-Blöbauma et al. 2013; Stern and Grinspun 2009).

1.4 Objectives and Layout of the Chapter

This chapter presents an introduction to the formulation of variational methods for finite-dimensional Lagrangian dynamical systems.

In Sect. 2 we revisit the Lagrangian and Hamiltonian points of view of mechanics for systems evolving on manifolds. To this end, we first introduce the concepts of generalized coordinates, configuration space, tangent space and the consistent computation of variations. Every concept is illustrated through some simple examples. Special attention is given to the construction of Lagrangian functions and to the role of Hamilton's variational principle in the deduction of the balance equations. In Sect. 2.3, we introduce the concept of (group) symmetries of the Lagrangian functional and we explain how they are related with the existence invariants of the dynamics. The symplectic nature of the flow is also discussed. Section 2.4 introduces the

[1] Alternative variational principles have been proposed for thermoelasticity with heat conduction, see e.g., (Yang et al. 2006; Vujanovic and Djukic 1971; Gambar and Markus 1994; Hutter and Tervoort 2007; Cannarozzi and Ubertini 2001). Structure-preserving methods may be consulted in (Armero and Simo 1992; Gross and Betsch 2006, 2007; Simo and Miehe 1992) to name only a few of them. Outstanding among the most recent approaches are the methods formulated by Romero (2009, 2010) which may be considered as energy-momentum methods applied to irreversible thermoelastic systems.

Legendre transform to compute the conjugated momenta along with the Hamiltonian form of the balance equations presented as a dynamical system evolving over the phase space.

The discrete version of Lagrangian mechanics is the main topic of Sect. 3. The standard methodology to construct variational integrators is explained in Sect. 3.1 and the position-momentum of the methods is given in Sect. 3.2. Section 3.3 is devoted to the implementation of the algorithms. The approximation properties and the geometric characteristics of the resulting discrete trajectories, including a discrete version of the symplecticity, are studied in Sect. 3.5.

Finally, in Sect. 4, we exemplify the usage of the variational methodology to construct structure-preserving methods for two problems of practical interest. First, we develop a VI based on the trapezoidal rule for a free-flying body that is able to undergo arbitrarily large rotations and displacements in space. The second example corresponds to the formulation of an explicit, second-order accurate variational integrator for the finite element discretization of geometrically exact rods.

This chapter is further complemented with the appendices A, B, and C.

2 Lagrangian and Hamiltonian Mechanics for Finite-Dimensional Systems

In the following, we will introduce (or review) the basic concepts in Lagrangian and Hamiltonian Mechanics. We shall make the abstractions concrete by applying them to two working examples, a particle in a hoop, and two particles joined by a rigid rod moving in a plane.

A particle in a hoop. Consider a particle of mass m that can slide without friction on a rigid circular hoop of radius R. The hoop is rigidly attached to an inertial frame, see Fig. 1.

Two particles joined by a rigid rod in a plane. Consider two particles of mass m joined by a rigid *massless* rod of length $2L$ which can freely move in $\mathbb{R}^2$.

As we shall see, these simple examples contain a lot of the concepts we discuss next.

2.1 Basic Concepts

For the first part of these notes we are going to consider mechanical systems for which all possible positions or *configurations* of the system can be identified with a finite-dimensional configuration space or, more generally, *configuration manifold Q* (see also Sect. A). The configuration manifold Q is a datum of the mechanical system, or at most, a modeling assumption. Prototypical systems of this type are multibody

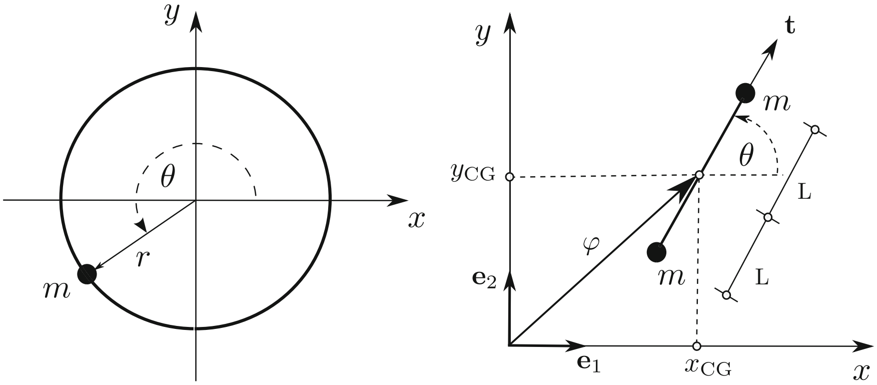


Fig. 1 Sketch of a particle in a hoop (*left*), and two particles joined by a rigid rod (*right*)

dynamical systems, involving a finite number of particles and rigid bodies moving together in $\mathbb{R}^n$, $n \in \mathbb{N}$.

To identify points in Q, we adopt a set of generalized coordinates

$$\mathbf{q} = (q_1, \ldots, q_d),$$

(lengths, angles, etc.), where $d \in \mathbb{N}$ is the dimension of Q. A *trajectory* of the mechanical system over the time interval $[0, T]$ is a map

$$\mathbf{q}(\cdot) \colon [0, T] \to Q.$$

In generalized coordinates, a trajectory is indicated with maps $q_i(t)$, $i = 1, \ldots, d$.

For mechanical systems consisting of m particles moving in some subset of $\mathbb{R}^k$, we can regard the configuration manifold as a subset of $\mathbb{R}^{k \times m}$. This means that we can define $k \times m$ maps $x_{\alpha,r}(q_1, \ldots, q_d)$, $r = 1, \ldots, k$, $\alpha = 1, \ldots, m$ that return the rth Cartesian coordinate of particle α for the given configuration of the system.

Particle in a hoop. $Q = S^1$, or the unit circle in $\mathbb{R}^2$. This is a one-dimensional manifold. The single generalized coordinate could be chosen as $q_1 = \theta$, the angle shown in Fig. 1. Notice that we could have chosen as a generalized coordinate any bijective and monotone function of θ, such as $q_1 = -\theta^3$. The choice of generalized coordinates is not unique. An example trajectory of the particle is $\theta(t) = \cos t$. For this system, we can define $x_{1,1}(\theta) = r \cos \theta$, $x_{1,2}(\theta) = r \sin \theta$, to recover the Cartesian coordinates of this particle in $\mathbb{R}^2$. Henceforth, we set $q_1 = \theta$ for this example.

Two particles joined by a rigid rod. $Q = \mathbb{R}^2 \times S^1$, which is a three-dimensional manifold, $d = 3$. A possible set of generalized coordinates is

$$(q_1, q_2, q_3) = (x_{CG}, y_{CG}, \theta),$$

where (x_{CG}, y_{CG}) are the Cartesian coordinates of the center of mass of the system and θ is the angle shown in Fig. 1. As before, other choices of generalized coordinates are possible, such as $(q_1, q_2, q_3) = (x_1, y_1, \theta)$, where (x_1, y_1) are the Cartesian coordinates of the position of one the masses. Henceforth, we choose the former set of generalized coordinates for this example. An example trajectory of the two particles is

$$(x_{CG}(t), y_{CG}(t), \theta(t)) = (t, 2t, t^2).$$

The Cartesian coordinates of this system in $\mathbb{R}^4$ are

$$\begin{aligned} x_{1,1}(x_{CG}, y_{CG}, \theta) &= x_{CG} + L\cos\theta \\ x_{1,2}(x_{CG}, y_{CG}, \theta) &= y_{CG} + L\sin\theta \\ x_{2,1}(x_{CG}, y_{CG}, \theta) &= x_{CG} - L\cos\theta \\ x_{2,2}(x_{CG}, y_{CG}, \theta) &= y_{CG} - L\sin\theta. \end{aligned}$$

□

Given a trajectory $\mathbf{q}(\cdot)$, the *generalized velocity* of the system at time t is $\dot{\mathbf{q}}(t)$. Given a point $\mathbf{q} \in Q$, the set of all possible generalized velocities of the system at $\mathbf{q}$ is called the *tangent space* $T_{\mathbf{q}}Q$, which is a vector space. The union of all points in Q with the tangent spaces attached is the *tangent bundle* TQ, and it is also a manifold. An element of the tangent bundle is, roughly speaking, a point $\mathbf{q} \in Q$ with a generalized velocity vector $\dot{\mathbf{q}}$ attached to it. Coordinates on TQ are denoted by

$$(\mathbf{q}, \dot{\mathbf{q}}) = (q_1, \dots, q_d, \dot{q}_1, \dots, \dot{q}_d).$$

For a system of m particles in $\mathbb{R}^k$, we can recover the traditional Cartesian components of the velocities as

$$v_{\alpha,r}(q_1, \dots, q_d, \dot{q}_1, \dots, \dot{q}_d) = \sum_{i=1}^{d} \dot{q}_i \frac{\partial x_{\alpha,r}}{\partial q_i}(q_1, \dots, q_d).$$

Particle in a hoop. $TQ = S^1 \times \mathbb{R}$. A point in TQ has coordinates $(\theta, \dot{\theta})$. For example, for the trajectory $\theta(t) = \cos t$, and the coordinates of the point of TQ in which the system is at $t = 1$ is $(\cos 1, -\sin 1)$. The tangent space at $\theta = \cos 1$, $T_{\cos 1}Q$, is the line tangent to the circle at such point, with origin at $\theta = \cos 1$. This is the space to which possible velocities of the particle at that point belong, so velocities are always tangent to the hoop. Figure 2 shows the position vector of the particle in space is given by

$$r^i = r\left(\cos\theta^i e_1 + \sin\theta^i e_2\right),$$

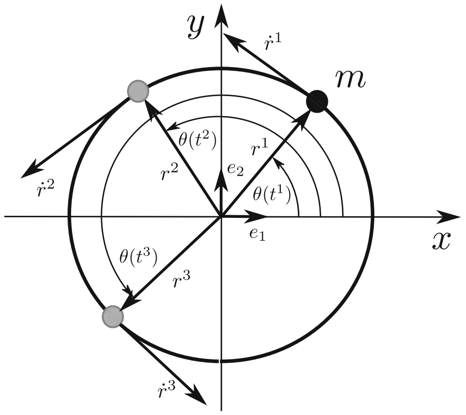


Fig. 2 Position vector, r^i, and velocity vector, $\dot{r}^i$, of the particle at times t^i ($i = 1, 2, 3$)

and the corresponding velocity vector by

$$\dot{r}^i = r\ \dot{\theta}^i \left(\cos\theta^i e_2 - \sin\theta^i e_1\right),$$

where θ^i and $\dot{\theta}^i$ denote the values of $\theta(t)$ and $\dot{\theta}(t)$ at the time t^i ($i = 1, 2, 3$), respectively. Note that we may think the motion of the particle in a hoop as a two-dimensional motion restricted by the condition $r^i \cdot \dot{r}^i = 0$.

Two particles joined by a rigid rod. $TQ = (S^1 \times \mathbb{R}^2) \times (\mathbb{R} \times \mathbb{R}^2)$, which has coordinates $(x_{CG}, y_{CG}, \theta, \dot{x}_{CG}, \dot{y}_{CG}, \dot{\theta})$. For the trajectory $(t, 2t, t^2)$, the generalized velocities at time $t = 1$ are $(1, 2, 2)$, and the coordinates in TQ are $(1, 2, 1, 1, 2, 2)$. A graphical depiction of the tangent space is difficult here, because we should be thinking about the tangent space to the surface defined by the configuration manifold when embedded in $\mathbb{R}^4$. □

Given $\mathbf{q}(\cdot)\colon [0, T] \to Q$, we consider a one-parameter family of trajectories $\mathbf{q}^\epsilon(\cdot) : [0, T] \to Q$ such that $\mathbf{q}_0(\cdot) = \mathbf{q}(\cdot)$ for all $\epsilon \in (-\varepsilon, \varepsilon)$, for some $\varepsilon > 0$. A *variation* $\delta\mathbf{q}(\cdot)$ of $\mathbf{q}(\cdot)$ is defined as

$$\delta\mathbf{q}(t) = \frac{d}{d\epsilon}\mathbf{q}^\epsilon(t)\Big|_{\epsilon=0}. \tag{1}$$

Coordinates of a variation are $(\delta q_1, \ldots, \delta q_d)$. Different one-parameter families of trajectories generally give rise to different variations, but of course, multiple one-parameter families of trajectories give rise to the same variation. Clearly, $(\mathbf{q}(t), \delta\mathbf{q}(t)) \in T_{\mathbf{q}(t)}Q$ for each t. An intuitive graphical interpretation of a variation is shown in Fig. 3.

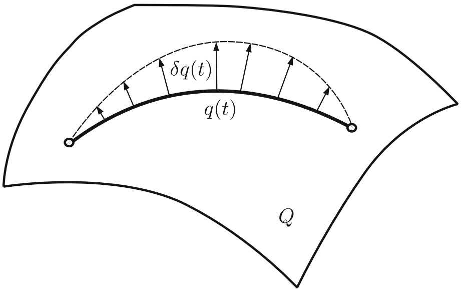


Fig. 3 Sketch of a trajectory $\mathbf{q}(\cdot)$ over Q, and a variation $\delta\mathbf{q}(\cdot)$. The variation at each time t, $\delta\mathbf{q}(t)$, is tangent to Q at $\mathbf{q}(t)$

For a system of m particles in $\mathbb{R}^k$, we can compute the Cartesian components of the variation as

$$\delta x_{\alpha,r} = \frac{d}{d\epsilon} x_{\alpha,r}(q_1^\epsilon, \ldots, q_d^\epsilon)\Big|_{\epsilon=0} = \sum_{i=1}^{d} \frac{\partial x_{\alpha,r}(q_1, \ldots, q_d)}{\partial q_i} \delta q_i.$$

Particle in a hoop. Consider the one-parameter family of trajectories $\theta^\epsilon(t) = \cos t + \epsilon \Delta\theta(t)$, for some $\Delta\theta : [0, T] \to \mathbb{R}$. Then, $\delta\theta(t) = \Delta\theta(t)$ at all times. The Cartesian components of the variation are

$$\delta x_{1,1}(t) = -r \sin\theta(t)\Delta\theta(t) \qquad \delta x_{1,2}(t) = r\cos\theta(t)\Delta\theta(t),$$

which clearly shows that $\delta q(t)$ is tangent to the circle at $q(t)$, for all times.

Another example of a variation follows by selecting $\theta^\epsilon(t) = \cos(t + \beta\epsilon)$, for $\beta \in \mathbb{R}$. In this case $\delta\theta(t) = -\beta \sin t$ at all times.

□

A *functional* is defined as a map from a set S to $\mathbb{R}$. Scalar-valued functions are functionals. More interesting functionals, however, are found when the set S contains functions, for example,

$$S[y(\cdot)] = \int_a^b y(t)\, dt \tag{2}$$

is a functional that takes values over the set S of integrable functions over $[a, b]$.

We are going to be interested in functionals that take values over sets of trajectories. The *variation of a functional* S at a trajectory $\mathbf{q}(\cdot)$ for a variation $\delta\mathbf{q}(\cdot)$ is defined as

$$\left\langle \delta S\left[\mathbf{q}(\cdot)\right], \delta\mathbf{q}\right\rangle = \frac{d}{d\epsilon} S\left[\mathbf{q}^\epsilon(\cdot)\right]\Big|_{\epsilon=0}, \tag{3}$$

where $\mathbf{q}^\epsilon(\cdot)$ is any of the one-parameter families that defines $\delta\mathbf{q}$. This is also called the Gâteaux derivative of S at $\mathbf{q}(\cdot)$ in direction $\delta\mathbf{q}(\cdot)$.

2.2 Lagrangian Mechanics

The starting point for Lagrangian mechanics is the definition of the *Lagrangian* $L : TQ \to \mathbb{R}$, $L(\mathbf{q}, \dot{\mathbf{q}})$, or in coordinates, $L(q_1, \dots, q_d, \dot{q}_1, \dots, \dot{q}_d)$. Notice that the Lagrangian returns a real number for each point of the tangent bundle TQ.

Different physical theories give rise to different Lagrangians. For mechanical systems, the Lagrangian has the general form $L = K - U$, where $K : TQ \to \mathbb{R}$ is the kinetic energy of the system, and $U : Q \to \mathbb{R}$ is the potential energy of the system.

Particle in a hoop. The Lagrangian for this system, in the presence of gravity in the negative-y direction is

$$L(\theta, \dot{\theta}) = \frac{m}{2} r^2 \dot{\theta}^2 - mgr \sin\theta. \tag{4}$$

Two particles joined by a rigid rod. In this case, in the absence of gravity,

$$L(x_{CG}, y_{CG}, \theta, \dot{x}_{CG}, \dot{y}_{CG}, \dot{\theta}) = m(\dot{x}_{CG}^2 + \dot{y}_{CG}^2 + L^2\dot{\theta}^2). \tag{5}$$

General multibody systems. A general class of Lagrangians obtained in multibody dynamics has the form

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} \dot{\mathbf{q}} \cdot \mathbf{M}(\mathbf{q}) \dot{\mathbf{q}} - U(\mathbf{q}), \tag{6}$$

where for each $\mathbf{q} \in Q$, $\mathbf{M}(\mathbf{q})$ is the symmetric and positive-definite $d \times d$ *mass matrix* of the system at $\mathbf{q}$. The particle in the hoop and the two particles connected by a rigid rod are particular cases of this Lagrangian.

Thermoelastic systems. Finite-dimensional and adiabatic thermoelastic systems may be constructed following (Maugin and Kalpakides 2002; Romero 2009; Mata and Lew 2011, 2012, 2014). We consider N masses connected by M thermoelastic springs, such as those shown in Fig. 4. The spatial position of the masses at time t is described by $\mathbf{q}(t) = (q_1(t), .., q_d(t)) \in Q_S$, where Q_S is the a d-dimensional manifold. Additionally, each thermoelastic spring is assigned a time-dependent *thermal displacement* $\Phi^i(t) \in \mathbb{R}$, $i = 1, \dots, M$, such that the empirical temperatures are computed as

$$\boldsymbol{\theta}(t) = \frac{d}{dt}\boldsymbol{\Phi}(t) = \left(\dot{\Phi}_1(t), \dots, \dot{\Phi}_M(t)\right).$$

The configuration manifold for this system is $Q = Q_S \times \mathbb{R}^M$, and it is specified by points of the form $(\mathbf{q}, \boldsymbol{\Phi})$. Trajectories of the system are time-dependent functions $(\mathbf{q}(t), \boldsymbol{\Phi}(t))$, and the generalized velocities are $(\dot{\mathbf{q}}(t), \boldsymbol{\theta}(t))$; therefore, in this system the temperatures of the springs are generalized velocities. The thermoelastic behavior of each spring is described by a Helmholtz free-energy

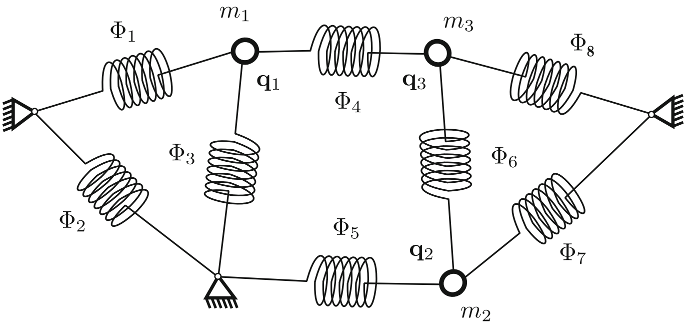


Fig. 4 A typical thermoelastic system: an assembly of three masses connected by eight thermoelastic springs and subjected to boundary conditions

function $\mathsf{A}_i(\mathbf{q}, \theta^i), i \in \{1, ..., M\}$ so that the Helmholtz free energy of the system follows as

$$\mathsf{A}(\mathbf{q}, \boldsymbol{\theta}) = \sum_{i=1}^{M} \mathsf{A}_i(\mathbf{q}, \theta^i).$$

A Lagrangian for this system is constructed as

$$L(\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\Phi}, \boldsymbol{\theta}) = \frac{1}{2}\dot{\mathbf{q}} \cdot \mathbf{M}(\mathbf{q})\dot{\mathbf{q}} - \mathsf{A}(\mathbf{q}, \boldsymbol{\theta}). \tag{7}$$

□

The second step in Lagrangian mechanics is the definition of the *action functional* over the time interval $[0, T]$

$$S\left[\mathbf{q}(\cdot)\right] = \int_0^T L\left(\mathbf{q}(t), \dot{\mathbf{q}}(t)\right)\, dt. \tag{8}$$

In Lagrangian mechanics the physical trajectories, namely, those that satisfy Newton's laws whenever the acceleration is well-defined, are obtained from a variational principle, *Hamilton's principle*. This principle states that: *The physical trajectory* $\mathbf{q}(\cdot)$ *is such that*

$$\left\langle S\left[\mathbf{q}(\cdot)\right], \delta\mathbf{q}\right\rangle = 0, \tag{9}$$

for all variations $\delta\mathbf{q}$ *that satisfy* $\delta\mathbf{q}(0) = \delta\mathbf{q}(T) = 0$. The set of all variations that satisfy these last conditions receive the name of *admissible variations*. Because the variation of S is the Gâteaux derivative of S, we also say that the action functional is stationary at the physical trajectory with respect to all admissible variations.

This variational principle completely characterizes trajectories of the system. It even gives meaning to physical trajectories when accelerations are not defined, and hence when Newton's second law cannot be applied. This is the case of impacts, for example. In principle, there would be no need to go any further to characterize trajectories. However, an alternative characterization of the stationary points of S is given by a system of ordinary differential equations that we shall find next. These are the equations of the trajectory of the system.

To find these equations, we proceed first by computing the variation of S for smooth enough $L\left(\mathbf{q}, \dot{\mathbf{q}}\right)$, $\mathbf{q}(\cdot)$, $\delta\mathbf{q}(\cdot)$:

$$\begin{aligned}
\langle \delta S[\mathbf{q}(\cdot)], \delta\mathbf{q}\rangle &= \frac{d}{d\epsilon} S[\mathbf{q}^{\epsilon}(\cdot)]\Big|_{\epsilon=0} \\
&= \frac{d}{d\epsilon}\int_0^T L(\mathbf{q}^{\epsilon}(t), \dot{\mathbf{q}}^{\epsilon}(t))\,dt\Big|_{\epsilon=0} \\
&= \sum_{i=1}^{d}\int_0^T \frac{\partial L}{\partial q_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\,\delta q_i(t) + \frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\,\delta \dot{q}_i(t)\,dt \\
&= \sum_{i=1}^{d}\int_0^T \left[\frac{\partial L}{\partial q_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\right)\right]\delta q_i(t)\,dt \\
&\qquad + \frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\,\delta q_i(t)\Big|_0^T. \qquad (10)
\end{aligned}$$

In Hamilton's principle $\delta\mathbf{q}(0) = \delta\mathbf{q}(T) = 0$, so the last term of the last expression is identically zero. Then, (9) implies that[2]

$$0 = \frac{\partial L}{\partial q_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\right) \qquad (11)$$

for all $t \in (0, T)$, and $i = 1, \ldots, d$. These are the Euler–Lagrange (E–L) equations of the system, and can be regarded as Newton's laws in terms of the chosen generalized coordinates. This is precisely part of the beauty of Lagrangian mechanics: The equations of motion are written as in (11) for *all* choices of generalized coordinates.

Particle in a hoop. The equations of motion are

$$0 = -mgr\cos\theta - \frac{d}{dt}\left(mr^2\dot{\theta}\right).$$

Two particles joined by a rigid rod. The equations of motion are:

$$m\frac{d}{dt}\dot{x}_{CG} = 0, \quad m\frac{d}{dt}\dot{y}_{CG} = 0 \quad \text{and} \quad mL^2\frac{d}{dt}\dot{\theta} = 0.$$

[2] The rigorous justification of this step requires the careful definition of the set of trajectories and corresponding variations, and then the use of some version of the fundamental lemma of the calculus of variations.

General multibody system. The equations of motion in this case take the form

$$0 = \frac{1}{2}\dot{\mathbf{q}} \cdot \frac{\partial \mathbf{M}}{\partial \mathbf{q}}(\mathbf{q})\dot{\mathbf{q}} - \frac{\partial U}{\partial \mathbf{q}}(\mathbf{q}) - \frac{d}{dt}\left(\mathbf{M}(\mathbf{q})\dot{\mathbf{q}}\right).$$

Thermoelastic systems. In this case, the equations of motion are given by

$$\begin{aligned}\frac{d}{dt}\left(\mathbf{M}(\mathbf{q})\dot{\mathbf{q}}\right) - \frac{1}{2}\dot{\mathbf{q}} \cdot \frac{\partial \mathbf{M}}{\partial \mathbf{q}}(\mathbf{q})\dot{\mathbf{q}} &= -\frac{\partial \mathsf{A}}{\partial \mathbf{q}}(\mathbf{q}, \boldsymbol{\theta}),\\ \frac{d}{dt}\left(-\frac{\partial \mathsf{A}}{\partial \boldsymbol{\theta}}(\mathbf{q}, \boldsymbol{\theta})\right) &= 0.\end{aligned}$$

To interpret these equations, it is useful to recall the thermodynamic relations (see, e.g., Coleman and Noll 1963),

$$\mathbf{f}(\mathbf{q}, \boldsymbol{\theta}) = -\frac{\partial \mathsf{A}}{\partial \mathbf{q}}(\mathbf{q}, \boldsymbol{\theta}) \quad \text{and} \quad \boldsymbol{\eta}(\mathbf{q}, \boldsymbol{\theta}) = -\frac{\partial \mathsf{A}}{\partial \boldsymbol{\theta}}(\mathbf{q}, \boldsymbol{\theta}), \tag{12}$$

where $\mathbf{f}$ is the vector of thermoelastic forces on the particles, and $\boldsymbol{\eta}$ is the vector of entropies, one entropy component per spring. Thus, the first equation states Newton's second law for the system in the case of a configuration-dependent mass matrix. The second equation states that the entropy of each one of the springs remains constant in time, as it should when no heat is transferred between springs. □

2.3 *Conservation Properties: Lagrangian Point of View*

A fundamental realization by E. Noether almost a century ago (Noether 1918) was that a (variation of a) motion that leaves the value of the Lagrangian invariant defines associated conserved quantities. Such motions are called symmetries of the Lagrangian, and we show some examples below. Typical examples of these conserved quantities are linear and angular momenta, and by adopting time as an independent coordinate as well, it is possible to regard energy as one such quantity as well (see, e.g., Kane et al. 1999). In summary, Noether's theorem states that to each symmetry of the Lagrangian corresponds a conserved quantity.

The simplest invariance or symmetry of the Lagrangian we may find is when for some i,

$$\frac{\partial L}{\partial q_i}(\mathbf{q}, \dot{\mathbf{q}}) = 0,$$

for all $(\mathbf{q}, \dot{\mathbf{q}}) \in TQ$. In this case a trajectory $\widehat{\mathbf{q}}$ that satisfies the E–L equations, satisfies that

$$\frac{\partial L}{\partial \dot{q}_i}(\widehat{\mathbf{q}}, \dot{\widehat{\mathbf{q}}})$$

is constant in time, and hence it is a conserved quantity for the motion $\widehat{\mathbf{q}}$. In such case, coordinate q_i is called a *cyclic coordinate*.

Of course, the definition of the symmetry is independent of the choice of coordinates. To keep the discussion at an intuitive level, it is useful to think about a symmetry of the Lagrangian in the following way: Given a trajectory $\mathbf{q}(t)$ and some $\varepsilon > 0$, we say that a one-parameter family of curves $\mathbf{q}^{\epsilon}(t)$ with $\mathbf{q}^{0}(t) = \mathbf{q}(t)$ is a symmetry of the Lagrangian if at all times t

$$L\left(\mathbf{q}^{\epsilon}(t), \dot{\mathbf{q}}^{\epsilon}(t)\right) = L\left(\mathbf{q}(t), \dot{\mathbf{q}}(t)\right) \tag{13}$$

for any $\epsilon \in (-\varepsilon, \varepsilon)$. The variation of this symmetry,

$$\boldsymbol{\xi}(t) = (\delta q_1(t), ..., \delta q_d(t))$$

computed according to (1), is called an *infinitesimal symmetry direction*, see e.g., (Marsden and West 2001; Lew et al. 2004).

If follows from integrating (13) in time that

$$S[\mathbf{q}^{\epsilon}(\cdot)] = S[\mathbf{q}(\cdot)] \tag{14}$$

for all $\epsilon \in (-\varepsilon, \varepsilon)$. Clearly, the variation of S in the infinitesimal symmetry direction is equal to zero, a result of computing the derivative with respect to ϵ on both sides of (14). Then, for a trajectory $\mathbf{q}(t)$ that satisfies the Euler–Lagrange equations (11) the only nonzero terms in (10) are the boundary terms, i.e.,

$$\begin{aligned} 0 &= \frac{d}{d\epsilon}\left(\int_0^T L(\mathbf{q}^{\epsilon}(t), \dot{\mathbf{q}}^{\epsilon}(t)dt\right)\Bigg|_{\epsilon=0} \\ &= \sum_{i=1}^{d}\left(\frac{\partial L}{\partial \dot{q}_i}\big(\mathbf{q}(T), \dot{\mathbf{q}}(T)\big)\, \xi_i(T) - \frac{\partial L}{\partial \dot{q}_i}\big(\mathbf{q}(0), \dot{\mathbf{q}}(0)\big)\, \xi_i(0)\right). \end{aligned} \tag{15}$$

The above equations are a formal statement of Noether's theorem which show that the initial and final values of the *momentum*

$$\mathbf{p}(t) = \left(\frac{\partial L}{\partial \dot{q}_1}\left(\mathbf{q}(t), \dot{\mathbf{q}}(t)\right), ..., \frac{\partial L}{\partial \dot{q}_d}\left(\mathbf{q}(t), \dot{\mathbf{q}}(t)\right)\right),$$

are equal in the $\boldsymbol{\xi}(t)$ direction,

$$\mathbf{p}(T) \cdot \boldsymbol{\xi}(T) = \mathbf{p}(0) \cdot \boldsymbol{\xi}(0). \tag{16}$$

The following examples illustrate the application of Noether's theorem.

Particle in a hoop. Consider the particle in the hoop in the *a*bsence of gravity and a one-parameter family of curves of the form $\theta^\epsilon(t) = \theta(t) + \epsilon\theta_0$ where θ_0 is a constant but otherwise arbitrary increment superposed onto $\theta(t)$. In words, we are considering trajectories identical to $\theta(t)$ that are simply rotated by a constant angle θ_0. A simple inspection reveals that the Lagrangian is invariant

$$L(\dot{\theta}^\epsilon) = L(\dot{\theta}),$$

and thus, according to Noether's theorem a conserved quantity exists. The infinitesimal symmetry direction in this case is

$$\delta\theta(t) = \theta_0,$$

and the momentum is

$$p_\theta(t) = \frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta}, \tag{17}$$

which is the angular momentum of the particle with respect to the center of the hoop.

From (16), we conclude that

$$p_\theta(T) = mr^2\dot{\theta}(T)\theta_0 = mr^2\dot{\theta}(0)\theta_0 = p_\theta(0),$$

for any θ_0, and hence that

$$mr^2\dot{\theta}(T) = mr^2\dot{\theta}(0),$$

which shows that in the absence of external potentials breaking the symmetry of the Lagrangian, the angular momentum is a constant of the motion.

Two particles joined by a rigid rod. For this example we consider the one-parameter family of curves

$$\begin{aligned}\mathbf{z}^\epsilon(t) &= \left(x^\epsilon_{CG}(t), y^\epsilon_{CG}(t), \theta^\epsilon(t)\right)\\ &= \left(x_{CG}(t) + \epsilon\bar{x}, y_{CG}(t) + \epsilon\bar{y}, \theta(t) + \epsilon\bar{\theta}\right)\\ &= \mathbf{z}(t) + \epsilon\, \boldsymbol{\chi},\end{aligned}$$

where $\boldsymbol{\chi} = (\bar{x}, \bar{y}, \bar{\theta})$ is an arbitrary vector in $\mathbb{R}^3$. The first two components of $\boldsymbol{\chi}$, $\mathbf{c} = (\bar{x}, \bar{y})$, represent an imposed rigid body translation in space, and $\bar{\theta}$ represents an imposed rigid body rotation, see Fig. 5. This family of curves is a symmetry of the Lagrangian, since it is simple to check that

$$L(\mathbf{z}^\epsilon, \dot{\mathbf{z}}^\epsilon) = L(\mathbf{z}, \dot{\mathbf{z}}).$$

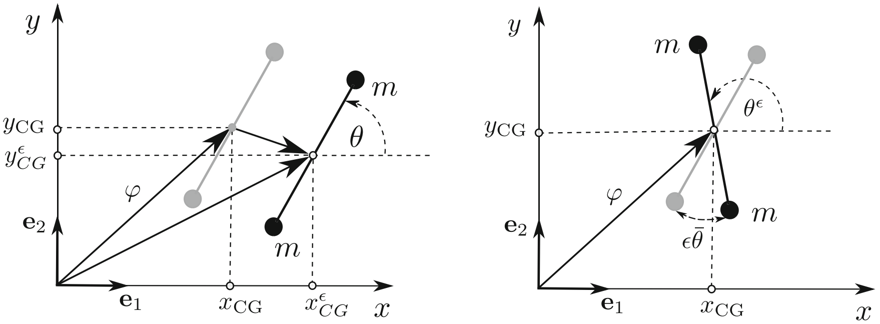


Fig. 5 Two particles joined by a rigid rod: rigid body translation (*left*), and rigid body rotation (*right*)

with infinitesimal symmetry direction

$$\delta\mathbf{z} = \frac{d}{d\epsilon}\mathbf{z}^{\epsilon}(t)\bigg|_{\epsilon=0} = \boldsymbol{\chi}.$$

The momentum in this case is computed as

$$\mathbf{p}(t) = 2m\left(\dot{x}_{CG}(t),\ \dot{y}_{CG}(t),\ L^2\dot{\theta}(t)\right).$$

It then follows from Noether's theorem that

$$\mathbf{p}(0)\cdot\boldsymbol{\chi} = \mathbf{p}(T)\cdot\boldsymbol{\chi}, \tag{18}$$

for any such $\boldsymbol{\chi}$. In particular, this implies that each component of $\mathbf{p}$ is conserved (choose $\boldsymbol{\chi} = (1, 0, 0)$, $\boldsymbol{\chi} = (0, 1, 0)$, and $\boldsymbol{\chi} = (0, 0, 1)$). Thus, the linear momentum of the system is constant in time

$$2m\dot{x}_{CG}(0) = 2m\dot{x}_{CG}(T), \qquad 2m\dot{y}_{CG}(0) = 2m\dot{y}_{CG}(T),$$

as it is the angular momentum of the system,

$$2mL^2\dot{\theta}(0) = 2mL^2\dot{\theta}(T).$$

Thermoelastic system. We consider the thermoelastic system described in Sect. 2.2 with constant mass matrix $\mathbf{M} = \mathrm{diag}(m, ..., m)$, $m \in \mathbb{R}$ and for which $Q_S \equiv \mathbb{R}^d$. For concreteness, we will let $\mathbf{q}_i$ denote the Cartesian coordinates of the ith particle in $\mathbb{R}^3$, and set $\mathbf{q} = (\mathbf{q}_1, \ldots, \mathbf{q}_N) = (q_1, \ldots, q_d) \in \mathbb{R}^d$. Moreover, we assume that a Helmholtz energy function of the form

$$\mathsf{A}(\mathbf{q},\boldsymbol{\theta}) = \sum_{i=1}^{M} \mathsf{A}_i\left(l_i(\mathbf{q}),\boldsymbol{\theta}\right),$$

where $l_i(\mathbf{q})$ denotes the distance between the two masses connected by the i^{th} thermoelastic spring.

We first consider a one-parameter family of curves $(\mathbf{q}^\epsilon, \boldsymbol{\Phi}^\epsilon)$ of the form

$$\begin{aligned}\mathbf{q}_i^\epsilon(t) &= \mathbf{q}_i(t) + \epsilon\mathbf{v} \qquad i = 1,\ldots,N\\ \boldsymbol{\Phi}^\epsilon(t) &= \boldsymbol{\Phi}(t) + \epsilon\mathbf{k},\end{aligned}$$

where $\mathbf{v} \in \mathbb{R}^3$ and $\mathbf{k} \in \mathbb{R}^M$ are constant but otherwise arbitrary vectors and $\epsilon \in \mathbb{R}$. This family of curves results from applying arbitrary rigid body displacements onto both the mechanical and the thermal positions of the system. As we see below, this is a symmetry of the Lagrangian, with infinitesimal symmetry direction

$$\boldsymbol{\xi} = (\mathbf{v}_N, \mathbf{k}),$$

where $\mathbf{v}_N = \underbrace{(\mathbf{v},\ldots,\mathbf{v})}_{\text{N times}} \in \mathbb{R}^d$. To check that this is a symmetry of the Lagrangian (7), note that

$$\dot{\mathbf{q}}^\epsilon(t) = \dot{\mathbf{q}}(t) \quad \text{and} \quad \dot{\boldsymbol{\Phi}}^\epsilon(t) = \boldsymbol{\theta}^\epsilon(t) = \boldsymbol{\theta}(t),$$

and therefore the kinetic energy is invariant upon changing ϵ, i.e.,

$$\frac{1}{2}\dot{\mathbf{q}}^\epsilon \cdot \mathbf{M}\dot{\mathbf{q}}^\epsilon = \frac{1}{2}\dot{\mathbf{q}} \cdot \mathbf{M}\dot{\mathbf{q}}.$$

Additionally, considering that the distance between masses is conserved by rigid body translations in space, i.e., $l_i(\mathbf{q}^\epsilon) = l_i(\mathbf{q})$, and that the constant translations of the thermal displacements do not change the temperature, as stated above, we have that

$$\mathsf{A}\left(\mathbf{q}^\epsilon,\boldsymbol{\theta}^\epsilon\right) = \mathsf{A}\left(\mathbf{q},\boldsymbol{\theta}\right),$$

from where it follows that the Lagrangian is invariant as well, and hence that $(\mathbf{q}^\epsilon, \boldsymbol{\Phi}^\epsilon)$ is one of its symmetries.

The momentum vector has components

$$\mathbf{p}(t) = \left(m\dot{\mathbf{q}}(t),\ \eta\big(\mathbf{q}(t),\boldsymbol{\theta}(t)\big)\right). \tag{19}$$

It follows from Noether's theorem that

$$\mathbf{p}(T)\cdot\boldsymbol{\xi} = \mathbf{p}(0)\cdot\boldsymbol{\xi}, \tag{20}$$

for any $\boldsymbol{\xi} = (\mathbf{v}_N, \mathbf{k}) \in \mathbb{R}^d \times \mathbb{R}^M$. Equivalently,

$$\left(\sum_{i=1}^{N} m\dot{\mathbf{q}}_i(T)\right) \cdot \mathbf{v} = \left(\sum_{i=1}^{N} m\dot{\mathbf{q}}_i(0)\right) \cdot \mathbf{v}, \tag{21}$$

Thus, the linear momentum of the system is conserved (we can, for example, choose $\mathbf{v} = \boldsymbol{e}_i$ for $i = 1, 2, 3$ to conclude this, where $\{\boldsymbol{e}_i\}_i$ is a basis in $\mathbb{R}^3$), namely,

$$\sum_{i=1}^{N} m\dot{\mathbf{q}}_i(T) = \sum_{i=1}^{N} m\dot{\mathbf{q}}_i(0). \tag{22}$$

Similarly, we can conclude that the entropy of each spring $\eta_i = -\partial A_i/\partial\theta^i$ is conserved (this follows by choosing $\boldsymbol{\xi}^j = (\mathbf{0}, \mathbf{k}^j)$ for $j = 1, \ldots, M$, where $\mathbf{k}_i^j = \delta_i^j$), namely,

$$\eta_i(0) = \eta_i(T),$$

for $i = 1, \ldots, M$. This is precisely what is expected from a system without heat conduction, and it follows as a consequence of the symmetry of the Lagrangian upon rigid translations of the thermal displacements.

Next, we consider a second family of one-parameter curves, which involve rigid rotations of the mechanical displacements and leave the thermal ones unaltered. The family of curves is

$$\begin{aligned} \mathbf{q}_i^\epsilon(t) &= \exp(\epsilon\widetilde{\boldsymbol{\omega}})\mathbf{q}_i(t), \qquad i = 1, ..., N, \\ \boldsymbol{\Phi}^\epsilon(t) &= \boldsymbol{\Phi}(t), \end{aligned}$$

where

$$\widetilde{\boldsymbol{\omega}} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} = \text{skew}[\widehat{\boldsymbol{\omega}}],$$

is a skew-symmetric but otherwise constant tensor, $\widehat{\boldsymbol{\omega}} = (\omega_1, \omega_2, \omega_3)$ is the axial vector of $\widetilde{\boldsymbol{\omega}}$, which satisfies $\widetilde{\boldsymbol{\omega}}\mathbf{v} = \widehat{\boldsymbol{\omega}} \times \mathbf{v}$ for any $\mathbf{v} \in \mathbb{R}^3$, and $\exp[\cdot]$ is the tensor exponential operator; see Sect. B for a brief introduction to finite rotations.

Because this family of curves rigidly rotates the trajectory, it is simple to verify that the magnitude of the velocity of each particle does not change with ϵ, and neither does the distance between any two particles. Therefore, this family of curves is also a symmetry of the Lagrangian, with infinitesimal symmetry direction given by

$$\boldsymbol{\xi}_Q(t) = \left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N, \mathbf{0}_M\right),$$

where $\mathbf{0}_M$ is an M-dimensional vector of zeroes and

$$\boldsymbol{\xi}_i(t) = \frac{d\mathbf{q}_i^\epsilon}{d\epsilon}(t)\bigg|_{\epsilon=0} = \widetilde{\boldsymbol{\omega}}\ \mathbf{q}_i(t) = \widehat{\boldsymbol{\omega}} \times \mathbf{q}_i(t), \quad i = 1, ..., N.$$

With this new infinitesimal symmetry direction in Noether's theorem (20) using the momentum (19) we obtain that

$$\sum_{i=1}^{N} \left(\mathbf{q}^i(T) \times m\dot{\mathbf{q}}^i(T)\right) \cdot \widehat{\boldsymbol{\omega}} = \sum_{i=1}^{N} \left(\mathbf{q}^i(0) \times m\dot{\mathbf{q}}^i(0)\right) \cdot \widehat{\boldsymbol{\omega}}.$$

Again, since this holds for any $\widehat{\boldsymbol{\omega}} \in \mathbb{R}^3$, we can conclude that the angular momentum of the system

$$\mathbf{A}(t) := \sum_{i=1}^{N} \mathbf{q}^i(t) \times m\dot{\mathbf{q}}^i(t)$$

is conserved, namely,

$$\mathbf{A}(0) = \mathbf{A}(T).$$

□

2.3.1 Conservation of Energy

When the Lagrangian is a convex function of the generalized velocities,[3] such as when it is a quadratic function of $\dot{\mathbf{q}}$, the energy of the system is defined as

$$E(\mathbf{q}, \dot{\mathbf{q}}) = \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}})\,. \tag{23}$$

This is the case for a general multibody system, whose Lagrangian is (6). In this case, the energy takes the form

$$E(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}\dot{\mathbf{q}} \cdot \mathbf{M}(\mathbf{q})\dot{\mathbf{q}} + U(\mathbf{q}). \tag{24}$$

The energy of the system is conserved along its solution trajectory. This can easily be seen by computing its time derivative and using the Euler–Lagrange equations. However, this result can also be obtained if we note that for an autonomous Lagrangian the following relation hods

[3] And as assumed here, the Lagrangian does not depend explicitly on time. Such Lagrangian is said to be autonomous.

$$\int_0^T L\left(\mathbf{q}(t), \dot{\mathbf{q}}(t)\right)\, dt = \int_{0+\epsilon}^{T+\epsilon} L\left(\mathbf{q}(s-\epsilon), \dot{\mathbf{q}}(s-\epsilon)\right)\, ds = I(\epsilon)$$

for any $\epsilon \in \mathbb{R}$. Equivalently, since the Lagrangian does not depend explicitly of time, this statement says that a trajectory can be translated uniformly in time without changing the value of the action. This is called a time-translation symmetry, and there is a way to frame this symmetry in the context of Noether's theorem, which we shall not pursue here (see, e.g., Marsden and West 2001). Differentiating this last expression, we get

$$\begin{aligned} 0 &= \left.\frac{dI(\epsilon)}{d\epsilon}\right|_{\epsilon=0} \\ &= \left[\int_{0+\epsilon}^{T+\epsilon} \frac{d}{d\epsilon} L\left(\mathbf{q}(s-\epsilon), \dot{\mathbf{q}}(s-\epsilon)\right) ds\right]_{\epsilon=0} + L\left(\mathbf{q}(t), \dot{\mathbf{q}}(t)\right)\Big|_0^T \\ &= -\int_0^T \frac{\partial L}{\partial \dot{\mathbf{q}}}\ddot{\mathbf{q}}\, dt - \int_0^T \frac{\partial L}{\partial \mathbf{q}}\dot{\mathbf{q}}\, dt + L(\mathbf{q}, \dot{\mathbf{q}})\Big|_0^T . \end{aligned}$$

Integrating by parts the first term of the right-hand side

$$\int_0^T \frac{\partial L}{\partial \dot{\mathbf{q}}}\ddot{\mathbf{q}}\, dt = \left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\dot{\mathbf{q}}\right)\Bigg|_0^T - \int_0^T \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right)\dot{\mathbf{q}}\, dt,$$

and replacing in the above equation yields

$$\left.\frac{dI(\epsilon)}{d\epsilon}\right|_{\epsilon=0} = \int_0^T \left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}}\right)\dot{\mathbf{q}}\, dt + \left[L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{\partial L}{\partial \dot{\mathbf{q}}}\dot{\mathbf{q}}\right]_0^T = 0.$$

The term under the integral vanish identically since $\mathbf{q}(t)$ is a solution trajectory of the system. The remaining boundary term is precisely a statement of the conservation of the energy, i.e.,

$$E(T) = E(0).$$

Therefore, the energy of mechanical systems described by autonomous Lagrangians is an invariant of the dynamics, and it is a result of the invariance of the action upon time-translating a trajectory in time.

2.3.2 Symplecticity: Lagrangian Point of View

Lagrangian mechanical systems also have another important conservation property: they conserve a skew-symmetric bilinear form known[4] as the *symplectic Lagrangian*

[4] We recall that a bilinear form on a vector space V is a mapping $\boldsymbol{w} : V \times V \to \mathbb{R}$ that is linear in both arguments. It is skew-symmetric if $\boldsymbol{w}(\boldsymbol{u}, \boldsymbol{v}) = -\boldsymbol{w}(\boldsymbol{v}, \boldsymbol{u})$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$.

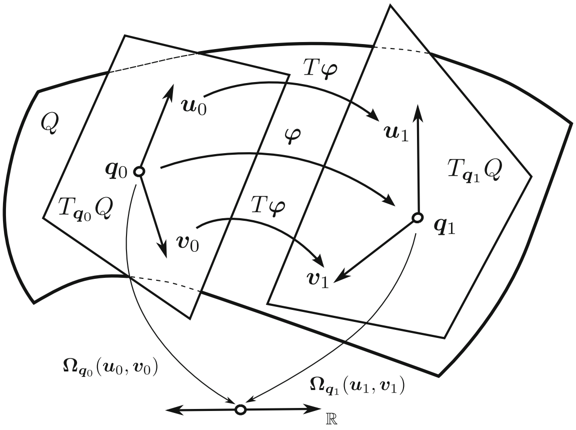


Fig. 6 Consider two points $\boldsymbol{q}_0, \boldsymbol{q}_1$ belonging to the configuration manifold Q such that $\boldsymbol{q}_1 = \varphi(\boldsymbol{q}_0)$. The tangent map $T\varphi$ relates the vectors $\boldsymbol{u}_0, \boldsymbol{v}_0 \in T_{\boldsymbol{q}_0}Q$ with $\boldsymbol{u}_1, \boldsymbol{v}_1 \in T_{\boldsymbol{q}_1}Q$ according to (25). The two-form $\boldsymbol{\Omega} : TQ \times TQ \to \mathbb{R}$ is conserved by φ if $\boldsymbol{\Omega}_{\boldsymbol{q}_0}(\boldsymbol{u}_0, \boldsymbol{v}_0) = \boldsymbol{\Omega}_{\boldsymbol{q}_1}(\boldsymbol{u}_1, \boldsymbol{v}_1)$

form along solution trajectories (Marsden and Ratiu 1999). In contrast to the conservation of energy or Noether's theorem, which are properties associated to individual trajectories, the symplectic form is associated to the behavior of nearby trajectories, namely, trajectories with very close initial conditions. We will discuss the closest to an "intuitive" explanation of the symplectic form we are aware of in a later section. We also refer the reader to Leimkuhler and Reich (2005), which contains a very approachable (the geometric concepts are progressively introduced) introduction and discussion on this topic.

Symplectic Map and Symplectic Form. We start by considering a N-dimensional configuration manifold Q with $N \in \mathbb{N}$ being an even number and a smooth enough and one-to-one mapping $\varphi : Q \to Q$. The image through the mapping φ of an arbitrary point $\boldsymbol{q}_0 \in Q$ is given by $\boldsymbol{q}_1 = \varphi(\boldsymbol{q}_0)$ which also belongs to Q. As explained in Sect. 2.1, it is possible to attach a tangent space to every point in a configuration manifold. Therefore, we construct the tangent spaces $T_{\boldsymbol{q}_0}Q$ and $T_{\boldsymbol{q}_1}Q$. See Fig. 6.

Additionally, we define *tangent map of* φ, denoted by $T\varphi$, as a mapping between elements of the tangent spaces according to

$$\begin{aligned} T\varphi : \quad & TQ && \to TQ \\ & (\boldsymbol{q}_0, \boldsymbol{u}_0) && \mapsto (\boldsymbol{q}_1, \boldsymbol{u}_1), \end{aligned}$$

where $\boldsymbol{q}_1 = \varphi(\boldsymbol{q}_0)$ and $\boldsymbol{u}_1 \in T_{\boldsymbol{q}_1}Q$ is obtained as

$$\boldsymbol{u}_1 := T\varphi \cdot \boldsymbol{u}_0 = \frac{d\varphi}{ds}(\boldsymbol{c}(s))\bigg|_{s=0}, \tag{25}$$

and $\boldsymbol{c}(s)$ is a s-parametrized curve on Q such that $\boldsymbol{c}'(0) = \boldsymbol{u}_0$. As it is schematically depicted in Fig. 6, $T\varphi$ maps elements in the tangent space of $\boldsymbol{q}_0$ to elements in the tangent space of $\boldsymbol{q}_1$. In terms of components, we have that

$$u_1^a = \sum_{j=1}^{N} \frac{\partial \varphi^a(\boldsymbol{q}_0)}{\partial q^j} u_0^j, \qquad a = 1, ..., N.$$

Further details can be found in (Marsden and Hughes 1983).

Furthermore, assume that we are given with a nondegenerated and skew-symmetric bilinear form defined for every $\boldsymbol{q} \in Q$ according to

$$\begin{aligned} \boldsymbol{\Omega}_{\boldsymbol{q}} : T_{\boldsymbol{q}} Q \times T_{\boldsymbol{q}} Q &\to \mathbb{R} \\ (\boldsymbol{u}, \boldsymbol{v}) \quad &\mapsto \Omega_{\boldsymbol{q}}(\boldsymbol{u}, \boldsymbol{v}). \end{aligned}$$

The explicit form of $\boldsymbol{\Omega}_{\boldsymbol{q}}$ is problem depend. Some examples are given in e.g., (Marsden and Ratiu 1999, Chap. 2). The subscript in $\boldsymbol{\Omega}_{\boldsymbol{q}}$ highlights the dependency of the two-form on the point of the manifold.

The map φ is called *symplectic* or *canonical* is it preserves $\boldsymbol{\Omega}_{\boldsymbol{q}}$ in the following sense:

$$\boldsymbol{\Omega}_{\boldsymbol{q}_0}(\boldsymbol{u}_0, \boldsymbol{v}_0) = \boldsymbol{\Omega}_{\varphi(\boldsymbol{q}_0)} \left(T\varphi \cdot \boldsymbol{u}_0, T\varphi \cdot \boldsymbol{v}_0\right),$$

where $\boldsymbol{u}_0, \boldsymbol{v}_0 \in T_{\boldsymbol{q}_0} Q$. Note that the above expression equals to zero, unless N is an even number. If this is the case, $\boldsymbol{\Omega}_{\boldsymbol{q}}$ is called a *symplectic two-form*. See Fig. 6 for a schematic representation.

In following, we explain how the solution trajectory of a Lagrangian mechanical system implicitly defines a time-dependent symplectic transformation over the tangent space of the configuration manifold, TQ. Furthermore, we show that the Lagrangian function of the system allows to construct the matrix representation of the corresponding symplectic two-form.

Lagrangian symplecticity. Consider a mechanical system characterized by a Lagrangian $L(\mathbf{q}, \dot{\mathbf{q}})$ evolving on a configuration manifold Q. We construct a solution trajectory over TQ by means considering the following smooth enough and one-to-one mapping,

$$\begin{aligned} \psi : TQ \times [0, T] &\to \quad TQ \\ (z, t) \quad &\mapsto \psi(z, t) \end{aligned} \tag{26}$$

where $z = (\mathbf{q}, \dot{\mathbf{q}})$ is an arbitrary point in TQ. The mapping ψ is such that $\psi(z, 0) = z$. Therefore, given certain initial conditions $z_0 \in TQ$, we define the trajectory of the mechanical system as

$$z(t) := \psi(z_0, t),$$

for all $t \in [0, T]$. Note that TQ posses the structure of a (even-dimensional) smooth manifold and therefore, it is possible to attach a tangent space to every $z \in TQ$. See Sect. A.1 and Fig. 7.

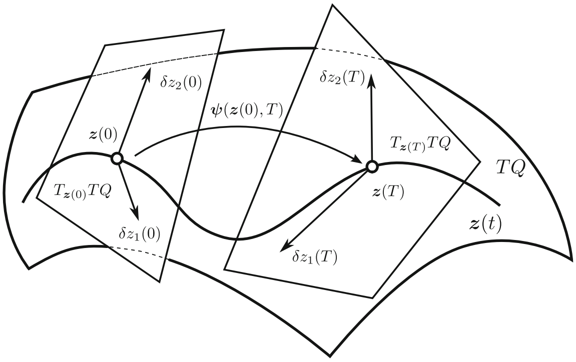


Fig. 7 Consider two points $z(0)$, $z(T)$ over the solution trajectory in the tangent space TQ. These points are related by ψ according to $z(T) = \psi(z(0), T)$. The variations $\delta z_i(0)$ and $\delta z_i(T)$, $i = 1, 2$, belong to the tangent spaces $T_{z(0)}TQ$ and $T_{z(T)}TQ$, respectively

A remarkable property of Lagrangian systems is given by the fact that $\psi(\cdot, t)$ results to be symplectic or in other words, there exist a skew-symmetric bilinear form, $\mathbf{\Omega}_L$, defined for every point over the solution trajectory $z(t)$ which enjoys the following conservation property,

$$\mathbf{\Omega}_L\big(z(0)\big)\Big(\delta z_1(0), \delta z_2(0)\Big) = \mathbf{\Omega}_L\big(z(T)\big)\Big(\delta z_1(T), \delta z_2(T)\Big), \tag{27a}$$

for arbitrary variations

$$\delta z_i(t) = \begin{bmatrix} \delta \mathbf{q}_i(t) \\ \delta \dot{\mathbf{q}}_i(t) \end{bmatrix} \in T_{z(t)}TQ, \quad i = 1, 2, \tag{27b}$$

such that

$$\delta z_i(T) = T\psi\big(z(0), T\big) \cdot \delta z_i(0), \qquad i = 1, 2. \tag{27c}$$

Moreover, the Lagrangian symplectic two-form may be represented in matrix form by

$$\mathbf{\Omega}_L(\mathbf{q}, \dot{\mathbf{q}}) = \begin{bmatrix} 0 & A_{12} & \dots & A_{1N} & B_{11} & \dots & B_{1N} \\ -A_{12} & 0 & \dots & A_{2N} & B_{21} & \dots & B_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -A_{1N} & -A_{2N} & \dots & 0 & B_{N1} & \dots & B_{NN} \\ -B_{11} & -B_{12} & \dots & -B_{1N} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -B_{N1} & -B_{N2} & \dots & -B_{NN} & 0 & \dots & 0 \end{bmatrix}, \tag{27d}$$

where

$$\begin{aligned} A_{ij}(\mathbf{q},\dot{\mathbf{q}}) &= -A_{ji}(\mathbf{q},\dot{\mathbf{q}}) = \frac{1}{2}\left(\frac{\partial^2 L}{\partial q_j \partial \dot{q}_i}(\mathbf{q},\dot{\mathbf{q}}) - \frac{\partial^2 L}{\partial q_i \partial \dot{q}_j}(\mathbf{q},\dot{\mathbf{q}})\right) \\ B_{ij}(\mathbf{q},\dot{\mathbf{q}}) &= \frac{\partial^2 L}{\partial \dot{q}_j \partial \dot{q}_i}(\mathbf{q},\dot{\mathbf{q}}). \end{aligned} \tag{27e}$$

To show how the above result can be deduced, we follow the procedure presented in (Marsden and West 2001; Lew et al. 2004). Consider a two-parameter family of solution trajectories $(\mathbf{q}^{\epsilon,\nu}(t), \boldsymbol{v}^{\epsilon,\nu}(t)) \in TQ$ with $\epsilon, \nu \in \mathbb{R}$ and compute the following variations

$$\delta\mathbf{q}_1^{\epsilon}(t) = \left.\frac{\partial \mathbf{q}^{\epsilon,\nu}}{\partial \nu}(t)\right|_{\nu=0} \tag{28a}$$

$$\delta\mathbf{q}_2^{\nu}(t) = \left.\frac{\partial \mathbf{q}^{\epsilon,\nu}}{\partial \epsilon}(t)\right|_{\epsilon=0} \tag{28b}$$

$$\delta^2\mathbf{q}(t) = \left.\frac{\partial^2 \mathbf{q}^{\epsilon,\nu}}{\partial \epsilon \partial \nu}(t)\right|_{\epsilon,\nu=0}, \tag{28c}$$

and the same applies for $\delta\boldsymbol{v}_1^{\epsilon}(t)$, $\delta\boldsymbol{v}_2^{\nu}(t)$, and $\delta^2\boldsymbol{v}(t)$. See Fig. 8. Moreover, we write

$$\delta\mathbf{q}_1(t) = \delta\mathbf{q}_1^0(t), \quad \delta\mathbf{q}_2(t) = \delta\mathbf{q}_2^0(t) \quad \text{and} \quad \mathbf{q}^{\epsilon}(t) = \mathbf{q}^{\epsilon,0}(t). \tag{28d}$$

Example. We build the two-parameters family of curves

$$\mathbf{q}^{\epsilon,\nu}(t) = \exp\left[\epsilon\widetilde{\boldsymbol{\omega}}\right](\mathbf{q}(t) + \nu\,\mathbf{k}),$$

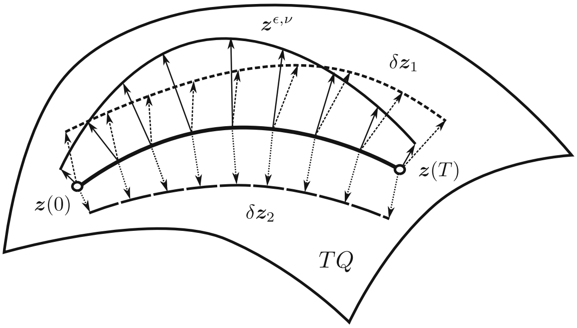


Fig. 8 Two-parameter family of solution trajectories $z(t)^{\epsilon,\nu} = (\boldsymbol{q}(t)^{\epsilon,\nu}, \dot{\mathbf{q}}(t)^{\epsilon,\nu}) \in TQ$. The corresponding variations $\delta z_1(t)$ and $\delta z_2(t)$ belong to the tangent space $T_{z(t)}TQ$

where $\widetilde{\boldsymbol{\omega}}$ is a constant and skew-symmetric tensor and $\mathbf{k} \in \mathbb{R}^3$. Then, we have

$$\begin{aligned}
\delta\mathbf{q}_1^{\epsilon}(t) &= \exp\left[\epsilon\widetilde{\boldsymbol{\omega}}\right]\mathbf{k}, \\
\delta\mathbf{q}_2^{\nu}(t) &= \widetilde{\boldsymbol{\omega}}\left(\mathbf{q}(t) + \nu\mathbf{k}\right), \\
\delta\mathbf{q}_1^{0}(t) &= \mathbf{k}, \\
\delta\mathbf{q}_2^{0}(t) &= \widetilde{\boldsymbol{\omega}}\mathbf{q}(t), \\
\delta^2\mathbf{q}(t) &= \left.\frac{d\delta\mathbf{q}_1^{\epsilon}}{d\epsilon}(t)\right|_{\epsilon=0} = \left.\frac{d\delta\mathbf{q}_2^{\nu}}{d\nu}(t)\right|_{\nu=0} = \widetilde{\boldsymbol{\omega}}\mathbf{k}.
\end{aligned}$$

□

Since we assumed that $\mathbf{q}^{\epsilon,\nu}(t)$ are solutions of the Euler–Lagrange equations, then

$$\left.\frac{\partial}{\partial\nu}\right|_{\nu=0} S\left[\mathbf{q}^{\epsilon,\nu}\right] = \sum_{i=1}^{N} \frac{\partial L}{\partial\dot{q}_i}(\mathbf{q}^{\epsilon,0}(t), \dot{q}^{\epsilon,0}(t))\ \delta\mathbf{q}_{1i}^{\epsilon}(t)\Big|_0^T \tag{29}$$

for any ϵ. We obtain the second variation of the action in the direction of these variations as (omitting arguments of functions for simplicity)

$$\begin{aligned}
\left.\frac{\partial}{\partial\epsilon}\right|_{\epsilon=0}\left.\frac{\partial}{\partial\nu}\right|_{\nu=0} S\left[\mathbf{q}^{\epsilon,\nu}\right] &= \left.\frac{\partial}{\partial\epsilon}\right|_{\epsilon=0}\left(\sum_{i=1}^{N}\frac{\partial L}{\partial\dot{q}_i}\ \delta\mathbf{q}_{1i}^{\epsilon}\Big|_0^T\right) \\
&= \sum_{i,j=1}^{N}\left[\frac{\partial^2 L}{\partial q_j\partial\dot{q}_i}\delta q_{1i}\delta q_{2j}\Big|_0^T + \frac{\partial^2 L}{\partial\dot{q}_j\partial\dot{q}_i}\delta q_{1i}\delta\dot{q}_{2j}\Big|_0^T\right] \\
&\quad + \sum_{i=1}^{N}\frac{\partial L}{\partial\dot{q}_i}\delta^2 q_i\Big|_0^T.
\end{aligned}$$

An equivalent expression is obtained by reversing the order of differentiation with respect to ϵ and ν since mixed partial derivatives are equal. Subtracting both expression we obtain

$$0 = \sum_{i,j=1}^{N}\left[\frac{\partial^2 L}{\partial q_j\partial\dot{q}_i}\left(\delta q_{1i}\delta q_{2j} - \delta q_{2i}\delta q_{1j}\right)\Big|_0^T + \frac{\partial^2 L}{\partial\dot{q}_j\partial\dot{q}_i}\left(\delta q_{1i}\delta\dot{q}_{2j} - \delta q_{2i}\delta\dot{q}_{1j}\right)\Big|_0^T\right]. \tag{30}$$

This identity can be rewritten to obtain an expression equivalent to (27a) with the $(2N) \times (2N)$ Lagrangian symplectic matrix given by (27d) and (27e), i.e.,

$$\left[\,\delta\mathbf{q}_1(0)\ \delta\dot{\mathbf{q}}_1(0)\,\right]^T\cdot\boldsymbol{\Omega}_L(\mathbf{q}(0),\dot{\mathbf{q}}(0))\cdot\begin{bmatrix}\delta\mathbf{q}_2(0)\\ \delta\dot{\mathbf{q}}_2(0)\end{bmatrix}=$$
$$\left[\,\delta\mathbf{q}_1(T)\ \delta\dot{\mathbf{q}}_1(T)\,\right]^T\cdot\boldsymbol{\Omega}_L(\mathbf{q}(T),\dot{\mathbf{q}}(T))\cdot\begin{bmatrix}\delta\mathbf{q}_2(T)\\ \delta\dot{\mathbf{q}}_2(T)\end{bmatrix}. \tag{31}$$

Therefore, we say that the Lagrangian symplectic two-form is exactly conserved along solution trajectories over TQ.

Particle in a hoop. Considering the Lagrangian function (4) we have that

$$\frac{\partial^2 L}{\partial\theta\partial\dot\theta}=0\quad\text{and}\quad\frac{\partial^2 L}{\partial\dot\theta^2}=mr^2,$$

and thus, the matrix representing the symplectic two-form is

$$\boldsymbol{\Omega}_L\big(\theta(t),\dot\theta(t)\big)=\begin{bmatrix}0 & mr^2\\ -mr^2 & 0\end{bmatrix}.$$

Consider, for example, the following bi-parametric family of solution trajectories

$$\begin{aligned}\theta^{\epsilon,\nu}(t)&=e^{\epsilon}\theta(t)+\nu t\\ \dot\theta^{\epsilon,\nu}(t)&=\frac{d\theta^{\epsilon,\nu}(t)}{dt}=e^{\epsilon}\dot\theta(t)+\nu,\end{aligned}$$

where $\epsilon,\nu\in\mathbb{R}$. Then,

$$\delta\theta_1(t)=t,\quad \delta\theta_2(t)=\theta(t),\quad \delta\dot\theta_1(t)=1\quad\text{and}\quad\delta\dot\theta_2(t)=\dot\theta(t).$$

Notice that $(\delta\theta_i(t),\delta\dot\theta_i(t))\in T_{\theta(t),\dot\theta(t)}TS^1$, $i=1,2$. The conservation of the symplectic two-form along solution trajectories (31) implies that

$$\begin{bmatrix}\delta\theta_1(0)\\ \delta\dot\theta_1(0)\end{bmatrix}\cdot\boldsymbol{\Omega}_L\begin{bmatrix}\delta\theta_2(0)\\ \delta\dot\theta_2(0)\end{bmatrix}=\begin{bmatrix}\delta\theta_1(T)\\ \delta\dot\theta_1(T)\end{bmatrix}\cdot\boldsymbol{\Omega}_L\begin{bmatrix}\delta\theta_2(T)\\ \delta\dot\theta_2(T)\end{bmatrix},$$

or equivalently,

$$\begin{bmatrix}0\\1\end{bmatrix}\cdot\begin{bmatrix}0 & mr^2\\ -mr^2 & 0\end{bmatrix}\begin{bmatrix}\theta(0)\\ \dot\theta(0)\end{bmatrix}=\begin{bmatrix}T\\1\end{bmatrix}\cdot\begin{bmatrix}0 & mr^2\\ -mr^2 & 0\end{bmatrix}\begin{bmatrix}\theta(T)\\ \dot\theta(T)\end{bmatrix}.$$

Two particles joined by a rigid rod. We note that the only nonzero coefficients of the Lagrangian symplectic two-form are

$$B_{11}=\frac{\partial^2 L}{\partial\dot x_{CG}^2}=2m,\quad B_{22}=\frac{\partial^2 L}{\partial\dot y_{CG}^2}=2m,\quad B_{33}=\frac{\partial^2 L}{\partial\dot\theta^2}=2mL^2,$$

and therefore, the symplectic two-form is represented by

$$\boldsymbol{\Omega}_L\big(z(t)\big) = \begin{bmatrix} \mathbf{0}_{3\times 3} & \mathbf{B} \\ -\mathbf{B} & \mathbf{0}_{3\times 3} \end{bmatrix},$$

where $z(t) = (x_{CG}(t), y_{CG}(t), \theta(t), \dot{x}_{CG}(t), \dot{y}_{CG}(t), \dot{\theta}(t)) \in TQ$, $\mathbf{0}_{3\times 3}$ is a 3×3 matrix with of zeros and

$$\mathbf{B} = 2m \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & L^2 \end{bmatrix}.$$

Consider two admissible variations $\delta z_i(t) \in T_{z(t)}TQ, i = 1, 2,$ computed according to (28a)–(28d). The conservation of the symplectic two-form along solution trajectories ensures that

$$\delta z_2(t) \cdot \boldsymbol{\Omega}_L\big(\mathbf{z}(t)\big)\delta z_1(t),$$

remains constant for all $t \in [0, T]$.

2.4 Hamiltonian Mechanics

Hamiltonian Mechanics reformulates Lagrange's equations of motion in generalized coordinates in a way that presents the motion of the system as a *flow* over *phase space*, as we explain next.

Given a Lagrangian L, the *conjugate momenta* are defined as

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}), \tag{32}$$

or in coordinates,

$$p_i = \frac{\partial L}{\partial \dot{q}_i}(\mathbf{q}, \dot{\mathbf{q}}) \qquad i = 1, \ldots, d. \tag{33}$$

This defines a map $(\mathbf{q}, \dot{\mathbf{q}}) \mapsto (\mathbf{q}, \mathbf{p})$, or $(\mathbf{q}, \mathbf{p}) = \mathrm{FL}(\mathbf{q}, \dot{\mathbf{q}})$. This map is termed the *Legendre transform*. The space of all possible values of $(\mathbf{q}, \mathbf{p})$ is called *phase space* T^*Q. To be precise, T^*Q is the cotangent bundle, also a manifold, which informally speaking is defined by attaching to each point $q \in Q$ the dual space T_q^*Q to T_qQ.

For typical mechanical systems $\mathrm{FL} : TQ \to T^*Q$ is bijective (and hence invertible) and onto, because the Lagrangian is strictly convex in $\dot{\mathbf{q}}$ for each $\mathbf{q}$. This means that for these systems all possible values of the conjugate momenta are attained at each point $\mathbf{q} \in Q$, and hence $\mathrm{FL}^{-1}(\mathbf{q}, \mathbf{p})$ is defined for all $\mathbf{p} \in \mathbb{R}^d$.

The *Hamiltonian* $H\colon T^*Q \to \mathbb{R}$ is defined as

$$H(\mathbf{p}, \mathbf{q}) = \mathbf{p} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}), \tag{34}$$

where $\dot{\mathbf{q}} = FL^{-1}(\mathbf{q}, \mathbf{p})$.

Lagrange's equations of motion can then be written in terms of $(\mathbf{q}, \mathbf{p})$ as (in coordinates)

$$\dot{q}_i = \frac{\partial H}{\partial p_i}(\mathbf{q}, \mathbf{p}) \tag{35a}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}(\mathbf{q}, \mathbf{p}). \tag{35b}$$

These are *Hamilton's equations of motion*, and are, again, valid for any choice of generalized coordinates. These equations follow easily from using (34), namely,

$$\frac{\partial H}{\partial p_j} = \dot{q}_j + \sum_{i=1}^{d} \left[p_i \frac{\partial \dot{q}_i}{\partial p_j} - \frac{\partial L}{\partial \dot{q}_i} \frac{\partial \dot{q}_i}{\partial p_j} \right] = \dot{q}_j$$

$$\frac{\partial H}{\partial q_j} = \sum_{i=1}^{d} \left[p_i \frac{\partial \dot{q}_i}{\partial q_j} - \frac{\partial L}{\partial q_j} - \frac{\partial L}{\partial \dot{q}_i} \frac{\partial \dot{q}_i}{\partial q_j} \right] = -\frac{\partial L}{\partial q_j}$$

and replacing in (11).

Particle in a hoop. The Legendre transform is

$$p_\theta = \mathrm{FL}(\theta, \dot{\theta}) = mr^2 \dot{\theta}$$

which is precisely the angular momentum of the particle around the origin. By solving the above equation for $\dot{\theta}$, and replacing in (34), we obtain the Hamiltonian

$$H(\theta, p_\theta) = \frac{p_\theta^2}{2mr^2} + mgr \sin\theta.$$

Hamilton's equations of motion are

$$\dot{\theta} = \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mr^2}$$

$$\dot{p}_\theta = -\frac{\partial H}{\partial \theta} = -mgr \cos\theta.$$

General multibody system. The Legendre transform in this case is

$$\boldsymbol{\pi} = \mathbf{M}(\mathbf{q})\dot{\mathbf{q}}. \tag{36}$$

Since $\mathbf{M}(\mathbf{q})$ is positive definite for any $\mathbf{q}$, it is invertible for any $\mathbf{q}$. Hence,

$$\dot{\mathbf{q}} = \mathbf{M}^{-1}(\mathbf{q})\boldsymbol{\pi}. \tag{37}$$

The Hamiltonian then follows as

$$H(\mathbf{q}, \boldsymbol{\pi}) = \frac{1}{2}\boldsymbol{\pi} \cdot \mathbf{M}^{-1}(\mathbf{q})\boldsymbol{\pi} + U(\mathbf{q}), \tag{38}$$

and Hamilton's equations of motion as

$$\begin{aligned} \dot{\mathbf{q}} &= \mathbf{M}^{-1}(\mathbf{q})\boldsymbol{\pi}, \\ \dot{\boldsymbol{\pi}} &= -\frac{1}{2}\boldsymbol{\pi} \cdot \frac{\partial \mathbf{M}^{-1}}{\partial \mathbf{q}}(\mathbf{q})\boldsymbol{\pi} - \frac{\partial U}{\partial \mathbf{q}}(\mathbf{q}). \end{aligned}$$

□

Hamilton's equations (35a) and (35b) can be written in a more succinct form as

$$\dot{\boldsymbol{y}} = \mathbf{J}\frac{\partial H}{\partial \boldsymbol{y}}(\boldsymbol{y}) = \boldsymbol{X}_H(\boldsymbol{y}), \tag{39}$$

where $\boldsymbol{y} = (\boldsymbol{q}, \boldsymbol{p})$ is a point on the phase space T^*Q, $\boldsymbol{X}_H(\boldsymbol{y}) \in T_{\boldsymbol{y}}T^*Q$ is known as the *Hamiltonian vector field* and the skew-symmetric matrix

$$\mathbf{J} = \begin{bmatrix} \mathbf{0}_{d\times d} & \mathbf{I}_{d\times d} \\ -\mathbf{I}_{d\times d} & \mathbf{0}_{d\times d} \end{bmatrix}, \tag{40}$$

is the *canonical symplectic matrix* which represents a skew-symmetric bilinear form, the *canonical symplectic two-fom*, that is conserved along solution trajectories on the phase space. This crucial aspect will be considered in detail in the following section. In (40) $\mathbf{I}_{d\times d}$ and $\mathbf{0}_{d\times d}$, are the $d \times d$ identity and zero matrices, respectively.

The expression (39) provides an intrinsic definition for Hamiltonian systems since given a smooth enough Hamiltonian function $H : T^*Q \to \mathbb{R}$, the matrix $\mathbf{J}$ may be though as a the linear transformation that maps $\partial H(\boldsymbol{y})/\partial \boldsymbol{y}$ which belongs to $T^*_{\boldsymbol{y}}T^*Q$, to the Hamiltonian vector field $\boldsymbol{X}_H(\boldsymbol{y})$ which belongs to $T_{\boldsymbol{y}}T^*Q$. See Fig. 9.

Note that as long as a solution of the system (39) exists, it is possible to define a function

$$\begin{aligned} \boldsymbol{\Phi} : T^*Q \times \mathbb{R} &\to \quad T^*Q \\ (\boldsymbol{y}, t) \quad &\mapsto \boldsymbol{\Phi}(\boldsymbol{y}, t) \end{aligned} \tag{41}$$

with $\boldsymbol{\Phi}(\boldsymbol{y}, 0) = \boldsymbol{y}_0$ so that the trajectory of the mechanical system that starts at $\boldsymbol{y}_0 \in T^*Q$ is given by

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}(\boldsymbol{y}_0, t). \tag{42}$$

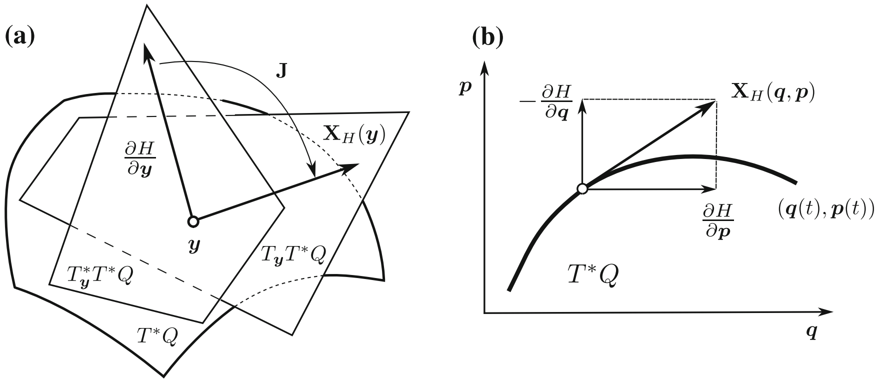

(a)
J
$\mathbf{X}_H(\boldsymbol{y})$
$\frac{\partial H}{\partial \boldsymbol{y}}$
$\boldsymbol{y}$
$T_{\boldsymbol{y}}T^*Q$
$T^*_{\boldsymbol{y}}T^*Q$
T^*Q
(b)
$\boldsymbol{p}$
$-\frac{\partial H}{\partial \boldsymbol{q}}$
$\mathbf{X}_H(\boldsymbol{q},\boldsymbol{p})$
$\frac{\partial H}{\partial \boldsymbol{p}}$
$(\boldsymbol{q}(t),\boldsymbol{p}(t))$
T^*Q
$\boldsymbol{q}$

Fig. 9 Consider a point $\boldsymbol{y} = (\mathbf{q}, \mathbf{p})$ belonging to the phase field T^*Q. **a** Shows how the gradient $\partial H(\boldsymbol{y})/\partial \boldsymbol{y}$ which belongs to the linear space $T^*_{\boldsymbol{y}}T^*Q$ is related to the Hamiltonian map $\mathbf{X}_H(\boldsymbol{y}) \in T_{\boldsymbol{y}}T^*Q$ through the map **J**. In (**b**) a geometric interpretation of Hamilton's equations is provided. The Hamiltonian vector field $\boldsymbol{X}_H(\boldsymbol{y}) = (\partial H/\partial \mathbf{p}, -\partial H/\partial \mathbf{q})$ is shown as a velocity vector which is tangent to the solution trajectory

The map $\boldsymbol{\Phi}$ is called the *flow map* of the Hamiltonian vector field $\mathbf{X}_H$, and it enables to consider the mechanical system as evolving collections of initial conditions, instead of only individual ones.

2.5 Conservation Properties: Hamiltonian Point of View

Since the Hamiltonian and Lagrangian points of view of mechanics are equivalent,[5] conservation properties may be stated in terms of variables on T^*Q. In this section we show how the properties described in Sect. 2.3 look when observed from the Hamiltonian point of view of mechanics. In particular, we present a comprehensive description of the connection existing between the symplectic nature of Hamiltonian flows and the conservation of volume in the phase space. This might seems at first glance, as a technicality of minor significance, however it has important consequences in the study of conservative and dissipative perturbations of nearly integrable Hamiltonian systems (Meyer et al. 2009; Maddocks and Overton 1995; Stoffer 1997, 1998; Hairer and Lubich 1999) and in the formulation of structure-preserving algorithms (Hairer et al. 2006, Chap. X, XI, XII; Channell and Scovel 1990; Meyer et al. 2011). The uninterested reader may skip this section, and continue directly to the formulation of variational integrators.

[5] There exist some exceptions in this regard such as for example, when the Legendre transform is not well defined.

2.5.1 Conservation of Energy

An autonomous Hamiltonian system conserves the energy, or in other words, its Hamiltonian function is exactly conserved along solution trajectories. This conservation property is readily verified by computing the time derivative of the Hamiltonian function (34) over a solution trajectory $(\mathbf{q}(t), \mathbf{p}(t))$ that satisfies (35a) and (35b), i.e.,

$$\dot{H}(\mathbf{q}, \mathbf{p}) = \frac{\partial H}{\partial \mathbf{p}} \cdot \dot{\mathbf{p}} + \frac{\partial H}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}} = -\frac{\partial H}{\partial \mathbf{p}} \cdot \frac{\partial H}{\partial \mathbf{q}} + \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{\partial H}{\partial \mathbf{p}} = 0. \tag{43}$$

2.5.2 Symplecticity of the Flow

An outstanding property of Hamiltonian systems is the *symplecticity* of their flows (41) on the phase space. In fact, the preservation of a discrete form of this property by the so-called *symplectic methods* (Yoshida 1993; Hairer et al. 2003) contributes to explain their superior performance in numerical simulations in terms of long-term stability and error propagation, when applied to Hamiltonian problems and/or to perturbations of them (Maddocks and Overton 1995; Hairer and Lubich 1999; Hairer et al. 2006, Chap. XII). A nice introduction to this subject can be consulted in (Leimkuhler and Reich 2005, Chap. 3) where a geometric interpretation of symplecticity in terms of volume preservation on the phase space is provided. A complete survey about symplectic forms on differentiable manifolds can be found elsewhere, e.g., (Arnold 1989; Marsden and Ratiu 1999).

To make a proper introduction of this geometric property of Hamiltonian flows, we closely follow (Leimkuhler and Reich 2005) and (Hairer et al. 2006). We first restrict the discussion to mappings from $\mathbb{R}^{2d}$ to itself and we show that for $d = 1$ the symplectic nature of the flow is manifested through area preservation on the phase space. The generalization to $d > 1$ allows to identify symplecticity with volume preservation. Finally, we provide some rudiments in order to extend the ideas to symplectic maps on cotangent bundles.

Symplectic two-fom. Consider two points $z_1 = (q_1^1, ..., q_1^n, p_1^1, ..., p_1^n)$ and $z_2 = (q_2^1, ..., q_2^n, p_2^1, ..., p_2^n)$ both belonging to $\mathbb{R}^{2n}$. The oriented area of the parallelogram spanned by the projections of z_1 and z_2 on the coordinate plane (q^i, p^i) is given by

$$\rho_i = p_2^i q_1^i - q_2^i p_1^i, \qquad i = 1, ...n, \tag{44}$$

see Fig. 10. The sum of all the oriented areas is

$$\omega\,(z_1, z_1) = \rho_1 + \cdots + \rho_n. \tag{45}$$

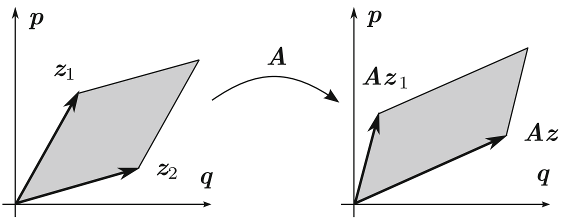


Fig. 10 Consider $z_1 = (q_1, p_1)$ and $z_2 = (q_2, p_2) \in \mathbb{R}^2$. The area of the parallelogram spanned is given by z_1 and z_2 is $\boldsymbol{\omega}(z_1, z^2) = p_1 q_2 - p_2 q_1$. The linear mapping $\boldsymbol{A} : \mathbb{R}^2 \to \mathbb{R}^2$ is symplectic if $\omega(z_1, z_2) = \omega(\mathbf{A}z_1, \mathbf{A}z_2)$. This is equivalent to area conservation

This expression defines the skew-symmetric bilinear map $\boldsymbol{\omega} : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \mapsto \mathbb{R}$ that may be represented in matrix form as

$$\boldsymbol{\omega}(z_1, z_2) = z_2 \cdot \mathbf{J}^{-1} z_1, \tag{46}$$

where $\mathbf{J}$ is the canonical symplectic matrix (40).

Symplectic mappings. A linear mapping $\boldsymbol{A} : \mathbb{R}^{2n} \mapsto \mathbb{R}^{2n}$ is called symplectic if

$$\boldsymbol{A}^{\mathrm{t}} \mathbf{J}^{-1} \boldsymbol{A} = \mathbf{J}^{-1}, \tag{47}$$

which is equivalent to

$$\boldsymbol{\omega}(z_1, z_2) = \boldsymbol{\omega}(\boldsymbol{A}z_1, \boldsymbol{A}z_2), \quad \text{for all } z_1, z_2 \in \mathbb{R}^{2n}. \tag{48}$$

Then, a symplectic linear map conserves the sum of the oriented areas of the parallelograms spanned by the projections of any two arbitrary points z_1 and z_2 on the coordinate planes $(q^i, p^i), i = 1, ..., n$. For $n = 1$ it is equivalent to area preservation in the phase space and for $n > 1$ to volume preservation. See Fig. 10.

For nonlinear maps, an analogous result holds. A differentiable map $\psi : U \subset \mathbb{R}^{2n} \mapsto \mathbb{R}^{2n}$ is symplectic if its Jacobian matrix $\partial\psi(\boldsymbol{y})/\partial\boldsymbol{y}$ is symplectic for all $\boldsymbol{y} \in U$, i.e.,

$$\left(\frac{\partial\psi}{\partial\boldsymbol{y}}(\boldsymbol{y})\right)^{\mathrm{t}} \mathbf{J}^{-1} \left(\frac{\partial\psi}{\partial\boldsymbol{y}}(\boldsymbol{y})\right) = \mathbf{J}^{-1}, \tag{49}$$

which is equivalent to (48) after replacing $\boldsymbol{A}$ by $\partial\psi(\boldsymbol{y})/\partial\boldsymbol{y}$. A geometric interpretation of symplecticity for nonlinear maps can be consulted in for example (Hairer et al. 2006, Chap. VI.2) or (Leimkuhler and Reich 2005).

Taking into account the above definitions it is possible to state the following fundamental result due to Poincaré:

Theorem (Poincaré 1899). For each fixed time t, the flow map $\boldsymbol{\Phi}(\boldsymbol{y}_0, t)$ of a Hamiltonian system with a twice continuously differentiable function $H(\boldsymbol{y})$, defines a symplectic transformation, i.e.,

$$\left(\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{y}_0}(\boldsymbol{y}_0, t)\right)^{\mathrm{t}} \mathbf{J}^{-1} \left(\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{y}_0}(\boldsymbol{y}_0, t)\right) = \mathbf{J}^{-1}. \tag{50}$$

Proof See (Hairer et al. 2006, pp. 184–185) or (Arnold 1989, Chap. 8) □

Equation (50) provides an intrinsic definition for Hamiltonian systems in the sense that any continuously differentiable $\boldsymbol{f} : U \subset \mathbb{R}^{2n} \mapsto \mathbb{R}^{2n}$ can be locally written as

$$\boldsymbol{f} = \mathbf{J}\frac{\partial H}{\partial \boldsymbol{y}}(\boldsymbol{y}),$$

for an appropriate Hamiltonian function $H(\boldsymbol{y})$ if the flow generated by

$$\dot{\boldsymbol{y}} = \boldsymbol{f}(\boldsymbol{y}),$$

is symplectic for all $\boldsymbol{y} \in U$ and small enough t (Hairer et al. 2006).

Hamiltonian flows on manifolds. The form in which Hamiltonian systems generate symplectic flows on the phase space can be alternatively explained following the ideas presented in Sect. 2.3.2. To this end, we consider a Hamiltonian system with differentiable flow map $\mathbf{\Phi}(\boldsymbol{y}, t)$ where $\boldsymbol{y} \in T^*Q$ and $t > 0$. We also consider the elements $\boldsymbol{u}$ and $\boldsymbol{v}$ which belong to $T_{\boldsymbol{y}}T^*Q$. Owing to the fact that $\mathbf{\Phi}$ is symplectic, it preserves the canonical symplectic form $\boldsymbol{\omega}$, according to

$$\begin{aligned}\boldsymbol{\omega}_{\boldsymbol{y}}\left(\boldsymbol{u}, \boldsymbol{v}\right) &= \boldsymbol{\omega}_{\mathbf{\Phi}(\boldsymbol{y},t)}\left(T\mathbf{\Phi}\cdot\boldsymbol{u}, T\mathbf{\Phi}\cdot\boldsymbol{v}\right),\\ \boldsymbol{u}\cdot\mathbf{J}^{-1}\boldsymbol{v} &= (T\mathbf{\Phi}\cdot\boldsymbol{u})\cdot\mathbf{J}^{-1}(T\mathbf{\Phi}\cdot\boldsymbol{v})\end{aligned}$$

which, since

$$T\mathbf{\Phi}(\boldsymbol{y}, t)\cdot\boldsymbol{u} = \frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{y}}(\boldsymbol{y}, t)\boldsymbol{u} = \mathbf{\Phi}'\boldsymbol{u},$$

yields to

$$\boldsymbol{u}\cdot\mathbf{J}^{-1}\boldsymbol{v} = \boldsymbol{u}\cdot\left(\mathbf{\Phi}'^{t}\mathbf{J}^{-1}\mathbf{\Phi}'\right)\boldsymbol{v}.$$

The condition given in (50) is recovered after noticing that $\boldsymbol{u}$ and $\boldsymbol{v}$ are arbitrary elements.

Figure 11 shows an illustration of the flow map Φ for the particle in the hoop setting $m = 1, r = 1$, and $g = 1$. The horizontal axis has the angle coordinate θ, and the vertical axis has the momentum, $mr^2\dot{\theta}$, which for these constants reduces to $\dot{\theta}$. We consider initial conditions belonging to the boundary of the square $[0, 4] \times [-2, 2] \in T^*Q \subset \mathbb{R}^2$ drawn in black color and to the boundary of the square $[0, 2] \times [-1, 1] \subset \mathbb{R}^2$ drawn in blue color. Then, we define the set of initial conditions as

$$\mathbf{B}^0 = \text{boundary}\big([0, 2] \times [-1, 1]\big) \,\cup\, \text{boundary}\big([0, 4] \times [-2, 2]\big). \tag{51}$$

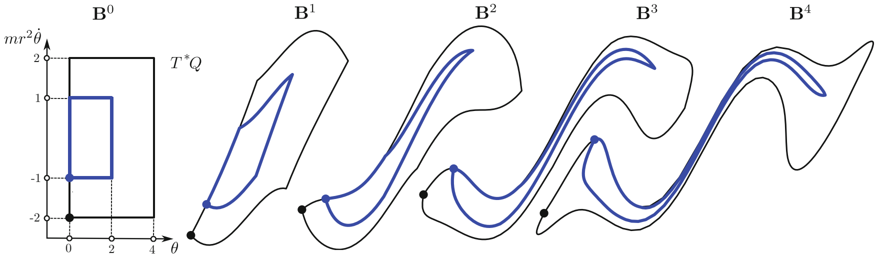


Fig. 11 Illustration of the flow map Φ for the particle in the hoop with $m = 1$, $r = 1$, and $g = 1$. The horizontal axis has the angle coordinate and the vertical axis has the momentum. For every time instant $\{t^i\}_{i=1,\ldots,4}$ the map Φ generates an updated configuration, $\mathbf{B}^i$, of the set of initial conditions (51) according to (52). Despite the fact that the shape of the set $\mathbf{B}^0$ deforms drastically in time, its area remains invariant due to the symplectic (volume-preserving) nature the flow

The area enclosed by the figure is 16. The flow map Φ deforms the shape of the set of initial conditions in such a manner that the sets

$$\mathbf{B}^i := \left\{ \boldsymbol{v} \in T^*Q \;\middle|\; \boldsymbol{v} = \Phi(\boldsymbol{u}_0, t^i),\ \boldsymbol{u}_0 \in \mathbf{B}^0 \right\}, \quad i = 1, 2, 3, 4. \tag{52}$$

represent updated configurations of $\mathbf{B}^0$ at later times $\{t^i\}_{i=1,\ldots,4}$. The same figure shows that even though the shape of the $\mathbf{B}^i$'s changes drastically, its area remains invariant due to the fact that Φ is a symplectic map.

3 Discrete Lagrangian Mechanics

This section focuses on presenting a systematic methodology to construct structure-preserving methods for finite-dimensional and autonomous, Lagrangian dynamical systems. To this end, we take advantage of some concepts in discrete Lagrangian mechanics (Marsden and West 2001). The basic idea is to consider discrete trajectories for the mechanical system, and to define the dynamics of those trajectories via a discrete version of Hamilton's principle applied to an approximate action.

3.1 Construction of Variational Integrators

We begin by considering a discretization $0 = t^0 < t^1 < \ldots < t^{N-1} < t^N = T$ of the interval $[0, T]$, for some $N \in \mathbb{N}$. In the most common case $t^n = nh$, with $h = T/N$, and for simplicity, this is the case we shall adopt for these notes. A *discrete trajectory* associated to this discretization is an element of

$$Q^{N+1} = \underbrace{Q \times \cdots \times Q}_{N \text{ times}},$$

and it is indicated by

$$\{\mathbf{q}^i\}_{i=0,\ldots,N} = \{\mathbf{q}^0, \mathbf{q}^1, \ldots, \mathbf{q}^{N-1}, \mathbf{q}^N\} \subset Q^{N+1}.$$

Let us then introduce a *discrete Lagrangian* $L_d \colon Q \times Q \times \mathbb{R} \to \mathbb{R}$ such that

$$L_d(\mathbf{q}^0, \mathbf{q}^1, h) \approx \int_0^h L(\mathbf{q}(t), \dot{\mathbf{q}}(t))\, dt \tag{53}$$

when $\mathbf{q}^0 = \mathbf{q}(0)$ and $\mathbf{q}^1 = \mathbf{q}(h)$, where $\mathbf{q}(t)$ is the exact trajectory of the system in that time interval. We shall discuss this approximation in more length later. Notice that the tangent space TQ has been replaced by $Q \times Q \times h$. Notice as well that the discrete Lagrangian is not an approximation of the Lagrangian, but rather of the action over a time interval $[0, h]$.

The discrete Lagrangian will be used to select a discrete trajectory for the system, inasmuch the Lagrangian is used to select a trajectory in the time-continuous system.

Example We explain how the discrete Lagrangians of some of the most popular time integrators for ODE's are constructed. We also apply the method to some particular Lagrangian functions.

- **Rectangle rule** (case 1). A general guideline for the construction of discrete Lagrangians is to combine an approximation space for a trajectory of the system and a quadrature rule for the integral over $[0, h]$. In this simplest case, we approximate the exact trajectory of the system $\mathbf{q}(t)$ over $[0, T]$ with a continuous piecewise linear polynomial, i.e.,

  $$\mathbf{q}(t) \approx \mathbf{q}^0 + \frac{\mathbf{q}^1 - \mathbf{q}^0}{h} t, \qquad t \in [0, h].$$

  Therefore, the exact velocity, $\dot{\mathbf{q}}(t)$, is approximated by

  $$h^{-1}(\mathbf{q}^1 - \mathbf{q}^0) \approx \dot{\mathbf{q}}(t), \qquad t \in [0, h],$$

  see Fig. 12. The discrete Lagrangian is then constructed as

  $$L_d^0(\mathbf{q}^0, \mathbf{q}^1, h) = hL\left(\mathbf{q}^0, \frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right) \approx \int_0^h L(\mathbf{q}(t), \dot{\mathbf{q}}(t))dt, \tag{54a}$$

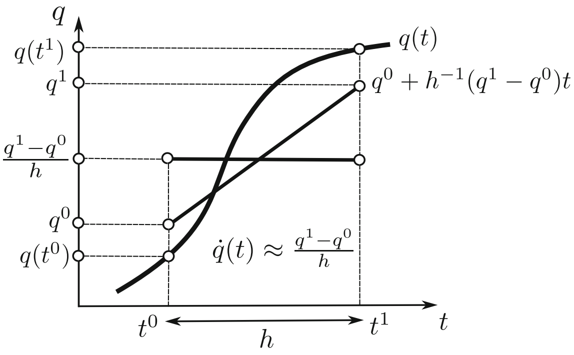


Fig. 12 Piecewise linear approximation of $\mathbf{q}(t)$, continuous across time intervals. The approximation obtained for the velocity is constant over the time interval, and hence generally discontinuous across time interval boundaries

which is equivalent to using a single quadrature point at the beginning of the interval [0, h] (rectangle rule). For the particle in the hoop, this discrete Lagrangian takes the form

$$L_d^0(\theta^0, \theta^1, h) = h\left[\frac{m}{2}r^2\left(\frac{\theta^1-\theta^0}{h}\right)^2 - mgr\sin\theta^0\right]. \tag{54b}$$

For thermoelastic systems we have

$$L_d^0(\mathbf{q}^0, \mathbf{q}^1, \boldsymbol{\Phi}^0, \boldsymbol{\Phi}^1, h) = \tag{54c}$$
$$h\left[\frac{1}{2}\left(\frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right)\cdot\mathbf{M}(\mathbf{q}^0)\left(\frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right) - \mathsf{A}\left(\mathbf{q}^0, \frac{\boldsymbol{\Phi}^1-\boldsymbol{\Phi}^0}{h}\right)\right].$$

- **Rectangle rule** (case 2). Alternatively, it is possible to approximate the action functional by

$$L_d^1(\mathbf{q}^0, \mathbf{q}^1, h) = hL\left(\mathbf{q}^1, \frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right). \tag{54d}$$

- **Trapezoidal rule**. Both, (54a) and (54d) are particular cases of the discrete Lagrangian

$$L_d^\alpha(\mathbf{q}^0, \mathbf{q}^1, h) = (1-\alpha)L_d^0(\mathbf{q}^0, \mathbf{q}^1, h) + \alpha L_d^1(\mathbf{q}^0, \mathbf{q}^1, h). \tag{54e}$$

which for $\alpha = \frac{1}{2}$ is known as the *trapezoidal rule*.
- **Midpoint rule**. The so-called *implicit midpoint method* is derived from the midpoint discrete Lagrangian

$$L_d^m(\mathbf{q}^0, \mathbf{q}^1, h) = hL\left(\frac{\mathbf{q}^1+\mathbf{q}^0}{2}, \frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right). \tag{54f}$$

For the two particles joined by the rigid rod, it is given by

$$\begin{aligned}&L_d(x^0, y^0, \theta^0, x^1, y^1, \theta^1, h) = \\ &\quad m\left[\frac{(x^1_{CG} - x^0_{CG})^2}{h} + \frac{(y^1_{CG} - y^0_{CG})^2}{h} + L^2\frac{(\theta^1_{CG} - \theta^0_{CG})^2}{h}\right],\end{aligned} \tag{54g}$$

and for thermoelastic systems,

$$\begin{aligned}&L^0_d(\mathbf{q}^0, \mathbf{q}^1, \mathbf{\Phi}^0, \mathbf{\Phi}^1, h) = \\ &\quad h\left[\frac{1}{2}\left(\frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right)\cdot \mathbf{M}\left(\frac{\mathbf{q}^1 + \mathbf{q}^0}{2}\right)\left(\frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right) - \right. \\ &\qquad \left. \mathsf{A}\left(\frac{\mathbf{q}^1 + \mathbf{q}^0}{2}, \frac{\mathbf{\Phi}^1 - \mathbf{\Phi}^0}{h}\right)\right].\end{aligned} \tag{54h}$$

Back to selecting the discrete trajectory, we construct the *discrete Action Sum* $S_d \colon Q^{N+1} \to \mathbb{R}$ as

$$S_d(\mathbf{q}^0, \dots, \mathbf{q}^N) = \sum_{i=0}^{N-1} L_d(\mathbf{q}^i, \mathbf{q}^{i+1}, h). \tag{55}$$

The *discrete Variational Principle* is formulated mimicking the continuous case: *The trajectory* $\{\mathbf{q}_i\}_{i=0,\dots,N}$ *is a stationary point of the discrete action sum among all variations that leave the endpoints fixed.* In other words,

$$\langle S_d(\mathbf{q}^0, \dots, \mathbf{q}^N), (\delta\mathbf{q}^0, \dots, \delta\mathbf{q}^N)\rangle = \sum_{i=0}^{N} \frac{\partial S_d}{\partial \mathbf{q}^i}(\mathbf{q}^0, \dots, \mathbf{q}^N)\delta\mathbf{q}^i = 0 \tag{56}$$

for any variations that satisfy $\delta\mathbf{q}^0 = \delta\mathbf{q}^N = 0$. This happens if and only if

$$\frac{\partial S_d}{\partial \mathbf{q}^i}\left(\mathbf{q}^0, \dots, \mathbf{q}^N\right) = 0, \qquad i = 1, \dots, N-1. \tag{57}$$

In terms of the discrete Lagrangians, this reads

$$0 = \frac{\partial L_d}{\partial \mathbf{q}^i}(\mathbf{q}^i, \mathbf{q}^{i+1}, h) + \frac{\partial L_d}{\partial \mathbf{q}^i}(\mathbf{q}^{i-1}, \mathbf{q}^i, h), \qquad i = 1, \dots, N-1. \tag{58}$$

These are the *discrete Euler–Lagrange equations*, or DEL equations, and they define the discrete trajectory. They also define the algorithm: If $\mathbf{q}^{i-1}$, $\mathbf{q}^i$ are known, these equations need to be solved to find $\mathbf{q}^{i+1}$.

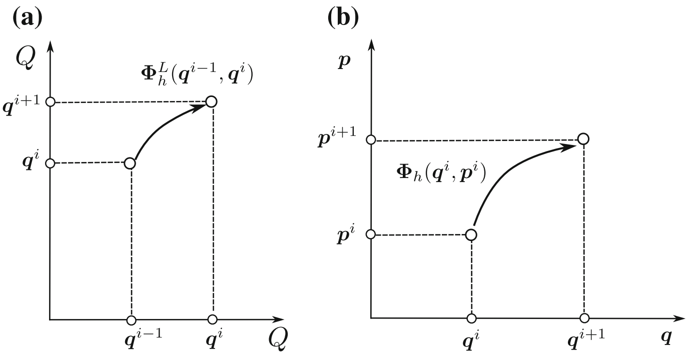


Fig. 13 Schematic representation of **a** the discrete Lagrangian map $\mathbf{\Phi}_h^L : Q \times Q \rightarrow Q \times Q$ and **b** the discrete Hamiltonian map $\mathbf{\Phi}_h : TQ^* \rightarrow TQ^*$

Thus, the discrete Euler–Lagrange equations implicitly define a map

$$\begin{array}{rl}\mathbf{\Phi}_h^L : & Q \times Q \quad \rightarrow Q \times Q \\ & (\mathbf{q}^{i-1}, \mathbf{q}^i) \mapsto \mathbf{\Phi}_h^L(\mathbf{q}^{i-1}, \mathbf{q}^i) = (\mathbf{q}^i, \mathbf{q}^{i+1})\end{array},$$

known as the *discrete Lagrangian map*. If $\partial L_d / \partial \mathbf{q}^i$ is invertible over $Q \times Q$, then this map is well defined. The map $\mathbf{\Phi}_h^L$ then flows the system forward over $Q \times Q$ between consecutive time-steps. See Fig. 13a.

As we shall see later, under simple conditions the discrete trajectories will approximate the exact trajectories of the Lagrangian in (53). The map $\mathbf{\Phi}_h^L$ can thus be considered a time-integrator, and because it satisfies the DEL equations of the discrete variational principle, it is called a *variational integrator*.

Example In this example, we derive the DEL equations for the discrete Lagrangians introduced so far, as applied to a general multibody system with a configuration-independent mass matrix.

- **Rectangle rules 1 and 2 and trapezoidal rule**. Consider first the discrete Lagrangian L_d^0 for such system. Then, we have that

$$\frac{\partial L_d^0}{\partial \mathbf{q}^0}(\mathbf{q}^0, \mathbf{q}^1, h) = -\mathbf{M}\left(\frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right) - h\frac{\partial U}{\partial \mathbf{q}}(\mathbf{q}^0)$$
$$\frac{\partial L_d}{\partial \mathbf{q}^1}(\mathbf{q}^0, \mathbf{q}^1, h) = \mathbf{M}\left(\frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right),$$

from where the DEL equations follow as

$$\begin{aligned} 0 &= -\,\mathbf{M}\left(\frac{\mathbf{q}^{i+1}-\mathbf{q}^{i}}{h}\right) - h\frac{\partial U}{\partial \mathbf{q}}(\mathbf{q}^{i}) + \mathbf{M}\left(\frac{\mathbf{q}^{i}-\mathbf{q}^{i-1}}{h}\right) \\ &= -\mathbf{M}\left(\frac{\mathbf{q}^{i+1}-2\mathbf{q}^{i}+\mathbf{q}^{i-1}}{h}\right) - h\frac{\partial U}{\partial \mathbf{q}}(\mathbf{q}^{i}). \end{aligned} \tag{59}$$

These are the equations of Newmark's second-order explicit algorithm, known also as central differences or Störmer-Verlet.

Since $\mathbf{M}$ is positive definite, it is possible to solve (59) to get

$$\mathbf{q}^{i+1} = 2\mathbf{q}^{i} - \mathbf{q}^{i-1} - h^2\mathbf{M}^{-1}\frac{\partial U}{\partial \mathbf{q}}(\mathbf{q}^{i}).$$

The discrete Lagrangian map is then

$$\Phi_h^L(\mathbf{q}^{i-1},\mathbf{q}^{i}) = \left(\mathbf{q}^{i}, 2\mathbf{q}^{i} - \mathbf{q}^{i-1} - h^2\mathbf{M}^{-1}\frac{\partial U}{\partial \mathbf{q}}(\mathbf{q}^{i})\right). \tag{60}$$

It is simple to check that the discrete Lagrangian

$$L_d^1(\mathbf{q}_0,\mathbf{q}_1,h) = hL\left(\mathbf{q}^1, \frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right),$$

gives rise to the same DEL in this case! Therefore, the *trapezoidal rule* (54e) yields the same DEL independently of the value chosen for α. This may not the case when the dependence of the Lagrangian on the velocities is more complex, such as in thermoelastic systems, see for example (Mata and Lew 2011, 2014).

- **Midpoint rule**. Using the midpoint rule, the discrete Lagrangian is given by

$$L_d^m(\mathbf{q}^0,\mathbf{q}^1,h) = h\left[\frac{1}{2}\left(\frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right)\cdot\mathbf{M}\left(\frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right) - U\left(\frac{\mathbf{q}^1+\mathbf{q}^0}{2}\right)\right], \tag{61}$$

and thus,

$$\begin{aligned} \frac{\partial L_d^m}{\partial \mathbf{q}^0}(\mathbf{q}^0,\mathbf{q}^1,h) &= -\mathbf{M}\left(\frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right) - \frac{h}{2}\frac{\partial U}{\partial \mathbf{q}}\left(\frac{\mathbf{q}^1+\mathbf{q}^0}{2}\right), \\ \frac{\partial L_d^m}{\partial \mathbf{q}^1}(\mathbf{q}^0,\mathbf{q}^1,h) &= \mathbf{M}\left(\frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right) - \frac{h}{2}\frac{\partial U}{\partial \mathbf{q}}\left(\frac{\mathbf{q}^1+\mathbf{q}^0}{2}\right). \end{aligned}$$

Therefore, the DEL equations are given by

$$
\begin{aligned}
0 = &-\mathbf{M}\left(\frac{\mathbf{q}^{i+1} - 2\mathbf{q}^{i} + \mathbf{q}^{i-1}}{h}\right) \\
&-\frac{h}{2}\left(\frac{\partial U}{\partial \mathbf{q}}\left(\frac{\mathbf{q}^{i+1} + \mathbf{q}^{i}}{2}\right) + \frac{\partial U}{\partial \mathbf{q}}\left(\frac{\mathbf{q}^{i} + \mathbf{q}^{i-1}}{2}\right)\right).
\end{aligned}
$$

Compare with (59). Note that in this case it is not possible to provide an explicit expression for the discrete Lagrangian map $\Phi_h^L(\mathbf{q}^{i-1}, \mathbf{q}^i)$. □

More Discrete Lagrangians. The discrete Lagrangians shown so far give rise to first- or second-order variational integrators. In the following, we show some discrete Lagrangians that give rise to higher order time integrators.

- **Quadratic rule**. Consider a piecewise continuous quadratic approximation of the trajectory over $[0, h]$, i.e.,

$$
\mathbf{q}^{t/h} = \mathrm{N}_1(t)\mathbf{q}^0 + \mathrm{N}_2(t)\mathbf{q}^{\frac{1}{2}} + \mathrm{N}_3(t)\mathbf{q}^1 \approx \mathbf{q}(t), \qquad t \in [0, h],
$$

where $\mathbf{q}^{\frac{1}{2}} \approx \mathbf{q}(h/2)$ is a coefficient needed for the quadratic interpolation and

$$
\mathrm{N}_i : [0, h] \to \mathbb{R}, \qquad i = 1, 2, 3,
$$

are a set of basis functions for the set of second degree polynomials over $[0, h]$, $\mathbb{P}^2([0, h])$, which satisfy

$$
\mathrm{N}_i(\tau_j) = \delta_{ij} \qquad \text{for } \tau_j \in \{0, h/2, h\},
$$

see Fig. 14a. The velocity is approximated by

$$
\dot{\mathbf{q}}^{t/h} = \dot{\mathrm{N}}_1(t)\mathbf{q}^0 + \dot{\mathrm{N}}_2(t)\mathbf{q}^{\frac{1}{2}} + \dot{\mathrm{N}}_3(t)\mathbf{q}^1 \approx \dot{\mathbf{q}}(t), \qquad t \in [0, h].
$$

The next step to construct a discrete Lagrangian consists in providing a quadrature rule, described by a set of quadrature points over $[0, h]$ and weights

$$
\{\xi_i, w_i\}_{i=1,\ldots,n_q}.
$$

Then, the discrete Lagrangian is given by

$$
L_d(\mathbf{q}^0, \mathbf{q}^1, h) = \inf_{\substack{\tilde{\mathbf{q}}(t)\,\in\,\mathbb{P}^2([0,h]) \\ \tilde{\mathbf{q}}(0)=\mathbf{q}^0,\ \tilde{\mathbf{q}}(h)=\mathbf{q}^1}} L_d^u(\mathbf{q}^0, \mathbf{q}^{1/2}, \mathbf{q}^1, h) \tag{62a}
$$

where the *unoptimized* Lagrangian is defined as

$$L_d^u(\mathbf{q}^0, \mathbf{q}^{1/2}, \mathbf{q}^1, h) = \sum_{i=1}^{n_q} w_i L(\widetilde{\mathbf{q}}(\xi_i), \dot{\widetilde{\mathbf{q}}}(\xi_i)). \tag{62b}$$

When a unique solution exists, the infimization in this last equation implicitly defines the value of $\mathbf{q}^{1/2}$ given those of $\mathbf{q}^0$ and $\mathbf{q}^1$. In practice, the infimization is imposed by requesting $\mathbf{q}^{1/2}$ to satisfy the stationarity condition

$$\frac{\partial L_d^u}{\partial \mathbf{q}^{1/2}} \left(\mathbf{q}^0, \mathbf{q}^{1/2}, \mathbf{q}^1, h\right) = \mathbf{0}. \tag{62c}$$

The DEL equations are then given by (58). In order to clarify the procedure consider the thermoelastic Lagrangian (7) with constant mass matrix along with the Gauss–Lobatto quadrature rule with points $\{0, h/2, h\}$ and weights $\{h/6, 4h/6, h/6\}$. First we define the *unoptimized* discrete Lagrangian

$$L_d^u(\mathbf{q}^0, \boldsymbol{\Phi}^0, \mathbf{q}^{\frac{1}{2}}, \boldsymbol{\Phi}^{\frac{1}{2}}, \mathbf{q}^1, \boldsymbol{\Phi}^{\frac{1}{2}}, h) = \frac{h}{6}\left(L^0 + 4L^{\frac{1}{2}} + L^1\right), \tag{62d}$$

where

$$\begin{aligned}
L^0 &= \dot{\mathbf{q}}^0 \cdot \mathbf{M}\dot{\mathbf{q}}^0 - \mathsf{A}\left(\mathbf{q}^0, \dot{\boldsymbol{\Phi}}^0\right),\\
L^{\frac{1}{2}} &= \dot{\mathbf{q}}^{\frac{1}{2}} \cdot \mathbf{M}\dot{\mathbf{q}}^{\frac{1}{2}} - \mathsf{A}\left(\mathbf{q}^{\frac{1}{2}}, \dot{\boldsymbol{\Phi}}^{\frac{1}{2}}\right),\\
L^1 &= \dot{\mathbf{q}}^1 \cdot \mathbf{M}\dot{\mathbf{q}}^1 - \mathsf{A}\left(\mathbf{q}^1, \dot{\boldsymbol{\Phi}}^1\right).
\end{aligned}$$

and

$$\begin{aligned}
\dot{\boldsymbol{\Phi}}^0 &= h^{-1}\left(4\boldsymbol{\Phi}^{\frac{1}{2}} - 3\boldsymbol{\Phi}^0 - \boldsymbol{\Phi}^1\right), &\quad \dot{\boldsymbol{\Phi}}^{\frac{1}{2}} &= h^{-1}\left(\boldsymbol{\Phi}^1 - \boldsymbol{\Phi}^0\right),\\
\dot{\boldsymbol{\Phi}}^1 &= h^{-1}\left(\boldsymbol{\Phi}^0 - 4\boldsymbol{\Phi}^{\frac{1}{2}} + 3\boldsymbol{\Phi}^1\right), &\quad \dot{\mathbf{q}}^0 &= h^{-1}\left(4\mathbf{q}^{\frac{1}{2}} - 3\mathbf{q}^0 - \mathbf{q}^1\right),\\
\dot{\mathbf{q}}^{\frac{1}{2}} &= h^{-1}\left(\mathbf{q}^1 - \mathbf{q}^0\right), &\quad \dot{\mathbf{q}}^1 &= h^{-1}\left(\mathbf{q}^0 - 4\mathbf{q}^{\frac{1}{2}} + 3\mathbf{q}^1\right).
\end{aligned}$$

Then, the discrete Lagrangian follows as

$$L_d(\mathbf{q}^0, \boldsymbol{\Phi}^0, \mathbf{q}^1, \boldsymbol{\Phi}^1, h) = \inf_{\mathbf{q}^{\frac{1}{2}}, \boldsymbol{\Phi}^{\frac{1}{2}}} L_d^u(\mathbf{q}^0, \boldsymbol{\Phi}^0, \mathbf{q}^{\frac{1}{2}}, \boldsymbol{\Phi}^{\frac{1}{2}}, \mathbf{q}^1, \boldsymbol{\Phi}^{\frac{1}{2}}, h). \tag{62e}$$

The values of $\mathbf{q}^{\frac{1}{2}}$ and $\boldsymbol{\Phi}^{\frac{1}{2}}$ that infimize L_d^u for given values of $\mathbf{q}^0$, $\boldsymbol{\Phi}^0$, $\mathbf{q}^1$ and $\boldsymbol{\Phi}^1$ satisfy the equations

$$\frac{\partial L_d^u}{\partial \mathbf{q}^{\frac{1}{2}}} = \frac{4\mathbf{m}}{h^2}(\mathbf{q}^k - 2\mathbf{q}^{k+\frac{1}{2}} + \mathbf{q}^{k+1}) - \mathbf{f}^{k+\frac{1}{2}} = \mathbf{0}, \tag{62f}$$

$$\frac{\partial L_d^u}{\partial \mathbf{\Phi}^{\frac{1}{2}}} = \mathbf{\Gamma}^{k+} - \mathbf{\Gamma}^{(k+1)-} = \mathbf{0}, \tag{62g}$$

where

$$\begin{aligned} \mathbf{f}^{k+\frac{1}{2}} &= -\frac{\partial \mathsf{A}}{\partial \mathbf{q}}\Big(\mathbf{q}^{k+\frac{1}{2}}, \frac{\mathbf{\Phi}^{k+1} - \mathbf{\Phi}^k}{h}\Big), \\ \mathbf{\Gamma}^{k+} &= -\frac{\partial \mathsf{A}}{\partial \theta}\Big(\mathbf{q}^k, \frac{-3\mathbf{\Phi}^k + 4\mathbf{\Phi}^{k+\frac{1}{2}} - \mathbf{\Phi}^{k+1}}{h}\Big), \\ \mathbf{\Gamma}^{k-} &= -\frac{\partial \mathsf{A}}{\partial \theta}\Big(\mathbf{q}^k, \frac{\mathbf{\Phi}^{k-1} - 4\mathbf{\Phi}^{k-\frac{1}{2}} + 3\mathbf{\Phi}^k}{h}\Big). \end{aligned}$$

- **General Galerkin variational integrators**. The basic idea behind the formulation of general Galerkin VI's consists in increasing the order of the approximation polynomial along with the order of the quadrature rule used to approximate the action of the system. To this end, we consider $s + 1$ control points over $[0, h]$,

  $$\{\mathbf{q}^{0_\nu}\}_{\nu=0,\ldots,s},$$

  which satisfy $\mathbf{q}^{0_0} = \mathbf{q}^0$ and $\mathbf{q}^{0_s} = \mathbf{q}^1$ and correspond to the values of the trajectory at control times

  $$\{d_\nu h\}_{\nu=0,\ldots,s} \quad \text{with} \quad 0 = d_0 < d_1 < \ldots < d_s = 1.$$

  Moreover, we assume that the exact trajectory over $[0, h]$ is approximated by a unique s-degree polynomial

  $$\mathbf{q}_p(t; \mathbf{q}^{0_0}, \ldots, \mathbf{q}^{0_s}, h) \in \mathbb{P}^s([0, h])$$

  such that

  $$\mathbf{q}_p(d_\nu h) = \mathbf{q}^{0_\nu}, \qquad \nu = 0, 1, \ldots, s.$$

  See Fig. 14b. Then, the discrete Lagrangian is obtained after providing an appropriate quadrature rule, $\{w_i, \xi_i\}_{i=1,\ldots,n_q}$, as

  $$L_d(\mathbf{q}^0, \mathbf{q}^1, h) = \inf_{\substack{\mathbf{q}(t)\in\mathbb{P}^s([0,h]) \\ \mathbf{q}(0)=\mathbf{q}^0, \mathbf{q}(h)=\mathbf{q}^1}} L_d^u\left(\mathbf{q}^{0_0}, \ldots, \mathbf{q}^{0_s}, h\right), \tag{62h}$$

  where

  $$L_d^u\left(\mathbf{q}^{0_0}, \ldots, \mathbf{q}^{0_s}, h\right) = \sum_{i=1}^{n_q} w_i L(\mathbf{q}_p(\xi_i), \dot{\mathbf{q}}_p(\xi_i)). \tag{62i}$$

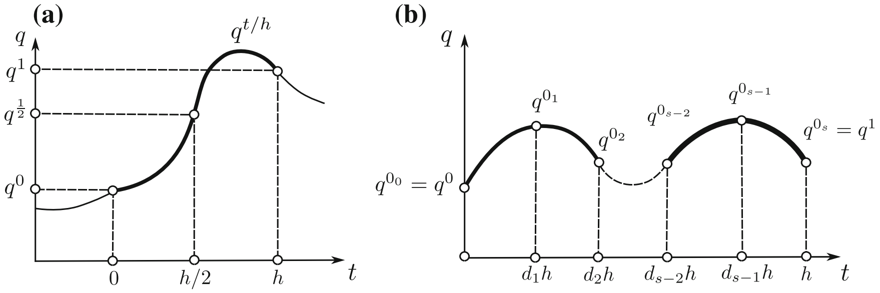


Fig. 14 **a** The polynomial $\mathbf{q}^{t/h}$ defines a continuous piecewise quadratic approximation of $\mathbf{q}(t)$ over $[0, h]$. **b** Higher order approximation of the trajectory by means of high-order polynomials

The infimization in (62h) implicitly defines the value of $\left\{\mathbf{q}^{0_\nu}\right\}_{\nu=1,\ldots,s-1}$ given those of $\mathbf{q}^0$ and $\mathbf{q}^1$. In this case the stationarity conditions are given by

$$\frac{\partial L_d^u}{\partial \mathbf{q}^{0_\nu}}\left(\mathbf{q}^{0_0}, \ldots, \mathbf{q}^{0_s}, h\right) = \mathbf{0}, \qquad \nu = 1, \ldots, s-1. \tag{63}$$

□

Example Consider the discrete Lagrangian (62e) for thermoelastic systems along with the stationarity (62f) and (62g). The DEL equations follow as

$$\frac{\mathbf{m}}{h^2}(-\mathbf{q}^{k+1} + 8\mathbf{q}^{k+\frac{1}{2}} - 14\mathbf{q}^k + 8\mathbf{q}^{k-\frac{1}{2}} - \mathbf{q}^{k-1}) = \frac{\mathbf{f}^{k+} + \mathbf{f}^{k-}}{2}, \tag{64a}$$

$$\boldsymbol{\Gamma}^{(k+1)-} - 4\boldsymbol{\Gamma}^{k+\frac{1}{2}} - 3\boldsymbol{\Gamma}^{k+} + 3\boldsymbol{\Gamma}^{k-} + 4\boldsymbol{\Gamma}^{k-\frac{1}{2}} - \boldsymbol{\Gamma}^{(k-1)+} = \mathbf{0}, \tag{64b}$$

where

$$\mathbf{f}^{k+} = -\frac{\partial \mathsf{A}}{\partial \mathbf{q}}\left(\mathbf{q}^k, \frac{-3\boldsymbol{\Phi}^k + 4\boldsymbol{\Phi}^{k+\frac{1}{2}} - \boldsymbol{\Phi}^{k+1}}{h}\right) \tag{65a}$$

$$\mathbf{f}^{k-} = -\frac{\partial \mathsf{A}}{\partial \mathbf{q}}\left(\mathbf{q}^k, \frac{\boldsymbol{\Phi}^{k-1} - 4\boldsymbol{\Phi}^{k-\frac{1}{2}} + 3\boldsymbol{\Phi}^k}{h}\right), \tag{65b}$$

$$\boldsymbol{\Gamma}^{k+\frac{1}{2}} = -\frac{\partial \mathsf{A}}{\partial \theta}\left(\mathbf{q}^{k+\frac{1}{2}}, \frac{\boldsymbol{\Phi}^{k+1} - \boldsymbol{\Phi}^k}{h}\right). \tag{65c}$$

Together, (62f), (62g), (64a) and (64b) provide enough equations to solve for $(\mathbf{q}^{k+\frac{1}{2}}, \mathbf{q}^{k+1}, \boldsymbol{\Phi}^{k+\frac{1}{2}}, \boldsymbol{\Phi}^{k+1})$ given $(\mathbf{q}^{k-\frac{1}{2}}, \mathbf{q}^k, \boldsymbol{\Phi}^{k-\frac{1}{2}}, \boldsymbol{\Phi}^k)$.

Remark Higher order methods can be constructed by means of dividing $[0, h]$ into subintervals and applying *composition methods*. The basic idea consists in combining several discrete Lagrangians together to obtain a new discrete Lagrangian with higher order of accuracy. In (Marsden and West 2001, pp. 49–51) the methodologies to construct both *multistep* and *single step, multisubsteps* methods are presented. □

3.2 Computation of Conjugate Momenta

The discrete Lagrangian map defines an evolution over $Q \times Q$, so it does not define either velocities or conjugate momenta. To do this, it is necessary to introduce the *discrete Legendre transforms* $\mathrm{F}^{\pm}L_d : Q \times Q \to T^*Q$ as

$$\mathrm{F}^{+}L_d(\mathbf{q}^0, \mathbf{q}^1, h) = (\mathbf{q}^1, \mathbf{p}^1) = \left(\mathbf{q}^1, \frac{\partial L_d}{\partial \mathbf{q}^1}(\mathbf{q}^0, \mathbf{q}^1, h)\right) \tag{66}$$

$$\mathrm{F}^{-}L_d(\mathbf{q}^0, \mathbf{q}^1, h) = (\mathbf{q}^0, \mathbf{p}^0) = \left(\mathbf{q}^0, -\frac{\partial L_d}{\partial \mathbf{q}^0}(\mathbf{q}^0, \mathbf{q}^1, h)\right). \tag{67}$$

So, the discrete Legendre transforms define conjugate momenta $\mathbf{p}^0$ and $\mathbf{p}^1$ at $\mathbf{q}^0$ and at $\mathbf{q}^1$, respectively. Notice that along a discrete trajectory a single value of a conjugate momentum is defined for each $\mathbf{q}^i$, instead of possibly two, coming each of the two surrounding time-steps. This is because the discrete trajectory satisfies the DEL equations, which state that

$$\frac{\partial L_d}{\partial \mathbf{q}^i}(\mathbf{q}^{i-1}, \mathbf{q}^i, h) = -\frac{\partial L_d}{\partial \mathbf{q}^i}(\mathbf{q}^i, \mathbf{q}^{i+1}, h), \tag{68}$$

so

$$\begin{aligned}\mathrm{F}^{-}L_d(\mathbf{q}^i, \mathbf{q}^{i+1}, h) &= \left(\mathbf{q}^i, -\frac{\partial L_d}{\partial \mathbf{q}^i}(\mathbf{q}^i, \mathbf{q}^{i+1}, h)\right)\\ &= \left(\mathbf{q}^i, \frac{\partial L_d}{\partial \mathbf{q}^i}(\mathbf{q}^{i-1}, \mathbf{q}^i, h)\right) = \mathrm{F}^{+}L_d(\mathbf{q}^{i-1}, \mathbf{q}^i, h)\end{aligned}$$

and the momentum at $\mathbf{q}^i$ is uniquely defined.

The introduction of the conjugate momentum permits rewriting the DEL in the so-called position-momentum form:

$$\mathbf{p}^i = -\frac{\partial L_d}{\partial \mathbf{q}^i}(\mathbf{q}^i, \mathbf{q}^{i+1}, h) \tag{69a}$$

$$\mathbf{p}^{i+1} = \frac{\partial L_d}{\partial \mathbf{q}^{i+1}}(\mathbf{q}^i, \mathbf{q}^{i+1}, h). \tag{69b}$$

Then, given $(\mathbf{q}^i, \mathbf{p}^i)$, these equations define $(\mathbf{q}^{i+1}, \mathbf{p}^{i+1})$. The map $\mathbf{\Phi}_h : T^*Q \to T^*Q$ defined as $(\mathbf{q}^{i+1}, \mathbf{p}^{i+1}) = \mathbf{\Phi}_h(\mathbf{q}^i, \mathbf{p}^i)$ is the *discrete Hamiltonian map*. See Fig. 13b.

The velocities do not have an intrinsic definition in discrete Lagrangian mechanics in terms of L_d, instead the velocities are approximated by inverting the Legendre transform, namely,

$$(\mathbf{q}^i, \dot{\mathbf{q}}^i) = FL^{-1}(\mathbf{q}^i, \mathbf{p}^i) \tag{70a}$$

or

$$\dot{\mathbf{q}}^i = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{q}^i, \mathbf{p}^i) \tag{70b}$$

for each i. Notice that $\dot{\mathbf{q}}^i$ defined in this way will generally be different than the approximation of the velocities that might have been used to construct the discrete Lagrangian. As we shall see later, approximating the velocity as in (70a) guarantees it will enjoy the same rate of convergence that $\mathbf{q}$ and $\mathbf{p}$ have.

Example In following we derive the position-momentum form of the DEL of some discrete Lagrangians.

- **Rectangle rules** (cases 1 and 2). The position-momentum form of the DEL corresponding to $L_d^0(\mathbf{q}^0, \mathbf{q}^1, h)$ is given by the equations

$$\begin{aligned}
\mathbf{p}^0 &= -\frac{\partial}{\partial \mathbf{q}^0}\left[hL\left(\mathbf{q}^0, \frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right)\right] &&= \left[-h\frac{\partial L}{\partial \mathbf{q}} + \frac{\partial L}{\partial \dot{\mathbf{q}}}\right]\Bigg|_{\left(\mathbf{q}^0, \frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right)}, \\
\mathbf{p}^1 &= \frac{\partial}{\partial \mathbf{q}^1}\left[hL\left(\mathbf{q}^0, \frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right)\right] &&= \frac{\partial L}{\partial \dot{\mathbf{q}}}\left(\mathbf{q}^0, \frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right),
\end{aligned} \tag{71}$$

  which for the case of a particle in the hoop are

$$\begin{aligned}
p^0 &= mr\left(r\left(\frac{\theta^1 - \theta^0}{h}\right) + hg\cos\theta^0\right), \\
p^1 &= mr^2\left(\frac{\theta^1 - \theta^0}{h}\right),
\end{aligned}$$

  therefore, the discrete Hamiltonian map is

$$(\theta^1, p^1) = \Phi_h^0(\theta^0, p^0) = \left(\theta^0 + \frac{hp^0}{mr^2} - \frac{h^2 g}{r}\cos\theta^0, p^0 - hmgr\cos\theta^0\right).$$

  For the case of $L_d^1(\mathbf{q}^0, \mathbf{q}^1, h)$ we have

$$\begin{aligned}
\mathbf{p}^0 &= -\frac{\partial}{\partial \mathbf{q}^0}\left[hL\left(\mathbf{q}^1, \frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right)\right] &&= \frac{\partial L}{\partial \dot{\mathbf{q}}}\left(\mathbf{q}^1, \frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right), \\
\mathbf{p}^1 &= \frac{\partial}{\partial \mathbf{q}^1}\left[hL\left(\mathbf{q}^1, \frac{\mathbf{q}^1 - \mathbf{q}^0}{h}\right)\right] &&= \left[h\frac{\partial L}{\partial \mathbf{q}} + \frac{\partial L}{\partial \dot{\mathbf{q}}}\right]\Bigg|_{\left(\mathbf{q}^1, \frac{\mathbf{q}^1-\mathbf{q}^0}{h}\right)}.
\end{aligned} \tag{72}$$

As we mentioned earlier, both discrete Lagrangians L_d^0 and L_d^1 lead to the same DEL. However, they define different approximations to the momenta, and hence to the velocity. This is how these two discrete Lagrangians, and hence their integrators, differ.

In the case of the particle in the hoop,

$$p^0 = mr^2 \left(\frac{\theta^1 - \theta^0}{h} \right),$$
$$p^1 = mr \left(r \left(\frac{\theta^1 - \theta^0}{h} \right) - hg \cos\theta^1 \right),$$

and the discrete Hamiltonian map is given by

$$(\theta^1, p^1) = \Phi_h^1(\theta^0, p^0) = \left(\theta^0 + \frac{hp^0}{mr^2}, p^0 - hmgr \cos\theta^1 \right).$$

The velocities are approximated considering (70a) as

$$\dot{\theta}^0 = FL^{-1}L(\theta^0, p^0) = \frac{p^0}{mr^2}$$
$$\dot{\theta}^1 = FL^{-1}(\theta^1, \dot{\theta}^1) = \frac{p^1}{mr^2}.$$

- **Trapezoidal rule**. If the trapezoidal rule is selected to construct the discrete Lagrangian for the particle in the hoop, $L_d^{\frac{1}{2}}(\theta^0, \theta^1, h)$, the following discrete Hamiltonian map is obtained

$$(\theta^1, p^1) = \Phi_h^{\frac{1}{2}}(\theta^0, p^0) = \frac{1}{2} \left(\Phi_h^0(\theta^0, p^0) + \Phi_h^1(\theta^0, p^0) \right)$$
$$= \left(\theta^0 + \frac{hp^0}{mr^2} - \frac{h^2 g}{2r} \cos\theta^0, p^0 - hmgr \frac{\cos\theta^1 - \cos\theta^0}{2} \right)$$

- **Midpoint rule**. Consider the midpoint discrete Lagrangian (54g). The position momentum form of the DEL equations is given by

$$\mathbf{p}^0 = \begin{pmatrix} p_x^0 \\ p_y^0 \\ p_\theta^0 \end{pmatrix} = -\frac{2m}{h} \begin{pmatrix} x_{CG}^1 - x_{CG}^0 \\ y_{CG}^1 - y_{CG}^0 \\ L^2(\theta^1 - \theta^0) \end{pmatrix}, \tag{73a}$$

and

$$\mathbf{p}^1 = \begin{pmatrix} p_x^1 \\ p_y^1 \\ p_\theta^1 \end{pmatrix} = -\mathbf{p}^0. \tag{73b}$$

□

The relations (69a) and (69b) are also valid for Galerkin VI's. However, they need to be complemented with (63) in order to fulfill the condition (62h).

Example Consider the discrete Lagrangian (62e) for the thermoelastic systems. The conjugate momenta are given by

$$\mathbf{p}^k = \frac{\mathbf{m}}{3h}(8\mathbf{q}^{k+\frac{1}{2}} - 7\mathbf{q}^k - \mathbf{q}^{k+1}) - \frac{h}{6}\mathbf{f}^{k+}, \tag{74a}$$

$$\mathbf{p}^{k+1} = \frac{\mathbf{m}}{3h}(\mathbf{q}^k - 8\mathbf{q}^{k+\frac{1}{2}} + 7\mathbf{q}^{k+1}) + \frac{h}{6}\mathbf{f}^{(k+1)-}, \tag{74b}$$

$$\eta^k = \frac{3}{6}\boldsymbol{\Gamma}^{k+} - \frac{1}{6}\boldsymbol{\Gamma}^{(k+1)-} + \frac{4}{6}\boldsymbol{\Gamma}^{k+\frac{1}{2}}, \tag{74c}$$

$$\eta^{k+1} = \frac{3}{6}\boldsymbol{\Gamma}^{(k+1)-} - \frac{1}{6}\boldsymbol{\Gamma}^{k+} + \frac{4}{6}\boldsymbol{\Gamma}^{k+\frac{1}{2}}. \tag{74d}$$

□

3.3 *Implementation of Variational Integrators*

The position-momentum form of the DEL equations provide a natural way to implement a variational integrator in a computer code. We describe the general format of an implementation next.

3.3.1 Initial Conditions

Most commonly initial conditions are provided in terms of positions and velocities instead of positions and momenta as required by variational integrators in position-momentum form. Then, we take advantage of the Legendre transform to compute the initial momentum as

$$\mathbf{p}^0 = \frac{\partial L}{\partial \dot{\mathbf{q}}}(\mathbf{q}(0), \dot{\mathbf{q}}(0)). \tag{75}$$

3.3.2 Basic Algorithm

The computer implementation of variational methods in position-momentum form follows a general structure summarized in Algorithm 1.

3.3.3 Post-processing Velocities

Consistent approximations to the velocities can be computed with the help of (35a) as

```
**Data**: Require $(\mathbf{q}^0, \mathbf{p}^0)$, $h$ and $N$.
**forall the** $k = 0, 1, \ldots, N-1$ **do**
    Solve $\mathbf{p}^k = -\frac{\partial L_d}{\partial \mathbf{q}^k}(\mathbf{q}^k, \mathbf{q}^{k+1}, h)$ for $\mathbf{q}^{n+1}$.
    Set $\mathbf{p}^{k+1} = \frac{\partial L_d}{\partial \mathbf{q}^{k+1}}(\mathbf{q}^k, \mathbf{q}^{k+1}, h)$.
**end**
```

Algorithm 1: Basic implementation of variational integrators.

$$\dot{\mathbf{q}}^k = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{q}^k, \mathbf{p}^k).$$

3.4 Approximation Properties and Convergence

In contrast to the traditional approach to constructing time integrators, which begins by approximating the equations of motion, the construction of a variational integrator departs from an approximation of the action. The order of convergence of a traditional integrator is generally assessed by the order of the consistency error. The question is then how the order of convergence of a variational integrator can be determined from the approximation of the action. This question was answered in (Marsden and West 2001, Sect. 2.3), and we describe the main ideas next.

First we define the *exact discrete Lagrangian*

$$L_d^E\left(\mathbf{q}^0, \mathbf{q}^1, h\right) = \int_0^h L\left(\mathbf{q}(t), \dot{\mathbf{q}}(t)\right)\, dt,$$

where $\mathbf{q}(t)$ is the solution of the E–L equations satisfying $\mathbf{q}(0) = \mathbf{q}^0$ and $\mathbf{q}(h) = \mathbf{q}^1$. In other words, L_d^E is a discrete Lagrangian that exactly matches the value of the action for the exact trajectory in the time interval $[0, h]$.

We can now define the *local variational order*. The discrete Lagrangian L_d is of order $r \geq 1$ if for any solution $\mathbf{q}$ of the E–L equations there exists $h_v > 0$ and $C_v > 0$ independent of h such that

$$\left\| L_d(\mathbf{q}(0), \mathbf{q}(h), h) - L_d^E(\mathbf{q}(0), \mathbf{q}(h), h) \right\| \leq C_v h^{r+1}, \tag{76}$$

for all $0 < h < h_v$.[6]

[6] Notice that C_v can depend on $(\mathbf{q}(0), \dot{\mathbf{q}}(0))$. For simplicity, we deliberately avoided the additional requirement that C_v should be uniformly bounded over a subset of TQ (see Marsden and West 2001). In designing a variational integrator and evaluating its order, this is a secondary condition not difficult to satisfy.

We are now ready to answer the question we started from: A fundamental result presented in (Marsden and West 2001, Theorem 2.3.1, pp. 43–44) states that *the variational integrator obtained from a discrete Lagrangian of order $r+1$ has order r*. Thus, to design a variational integrator of order r it is enough to construct a discrete Lagrangian of order $r+1$. We show examples of order calculation below.

3.4.1 Order Calculation

To compute the order of $L_d(\mathbf{q}(0), \mathbf{q}(h), h)$, we expand it in a Taylor series of h around $h = 0$ and compare the terms with those of the Taylor series expansion of the exact discrete Lagrangian. The first few terms of the latter are

$$\begin{aligned} L_d^E(\mathbf{q}(0), \mathbf{q}(h), h) &= hL\left(\mathbf{q}(0), \dot{\mathbf{q}}(0)\right) + \\ &+\frac{h^2}{2}\left(\frac{\partial L}{\partial \mathbf{q}}\left(\mathbf{q}(0), \dot{\mathbf{q}}(0)\right)\cdot\dot{\mathbf{q}}(0) + \frac{\partial L}{\partial \dot{\mathbf{q}}}\left(\mathbf{q}(0), \dot{\mathbf{q}}(0)\right)\cdot\ddot{\mathbf{q}}(0)\right) + \mathcal{O}(h^3). \end{aligned} \tag{77}$$

If the first r terms of the series of both discrete Lagrangian are the same, then the discrete Lagrangian is of order $r+1$.

Example Consider the discrete Lagrangian built on the trapezoidal rule,

$$\begin{aligned} L_d^\alpha(\mathbf{q}(0), \mathbf{q}(h), h) &= f(h) = \\ &(1-\alpha)hL\left(\mathbf{q}(0), \frac{\mathbf{q}(h)-\mathbf{q}(0)}{h}\right) + \alpha hL\left(\mathbf{q}(h), \frac{\mathbf{q}(h)-\mathbf{q}(0)}{h}\right). \end{aligned}$$

Here we introduced the name $f(h)$ to explicitly indicate that the left-hand side is a function of h only. Then,

$$\begin{aligned} f(0) &= L_d^\alpha(\mathbf{q}(0), \mathbf{q}(0), 0) &&= 0, \\ f'(0) &= \frac{dL_d^\alpha}{dh}(\mathbf{q}(0), \mathbf{q}(0), 0) &&= L(\mathbf{q}(0), \dot{\mathbf{q}}(0)), \\ f''(0) &= \frac{d^2L_d^\alpha}{dh^2}(\mathbf{q}(0), \mathbf{q}(0), 0) &&= 2\alpha\frac{\partial}{\partial \mathbf{q}}L(\mathbf{q}(0), \dot{\mathbf{q}}(0))\cdot\dot{\mathbf{q}}(0) + \frac{\partial}{\partial \dot{\mathbf{q}}}L(\mathbf{q}(0), \dot{\mathbf{q}}(0))\cdot\ddot{\mathbf{q}}(0), \end{aligned}$$

therefore,

$$\begin{aligned} L_d^\alpha(\mathbf{q}(0), \mathbf{q}(h), h) &= hL(\mathbf{q}(0), \dot{\mathbf{q}}(0)) \\ &\quad + \frac{h^2}{2}\left[2\alpha\frac{\partial}{\partial \mathbf{q}}L(\mathbf{q}(0), \dot{\mathbf{q}}(0))\cdot\dot{\mathbf{q}}(0)\right. \\ &\qquad\left. + \frac{\partial}{\partial \dot{\mathbf{q}}}L(\mathbf{q}(0), \dot{\mathbf{q}}(0))\cdot\ddot{\mathbf{q}}(0)\right] + \mathcal{O}\left(h^3\right), \end{aligned} \tag{78}$$

Comparing with (77) have that

$$L_d^E(\mathbf{q}(0),\mathbf{q}(h),h) - L_d^\alpha(\mathbf{q}(0),\mathbf{q}(h),h) = \frac{h^2}{2}(1-2\alpha)\frac{\partial}{\partial \mathbf{q}}L(\mathbf{q}(0),\dot{\mathbf{q}}(0))\cdot\dot{\mathbf{q}}(0) + \mathcal{O}\left(h^3\right),$$

and hence $r + 1 = 3$ if and only if $\alpha = 1/2$, and $r + 1 = 2$ otherwise. This means that the two rectangle rules ($\alpha = 0, 1$) are only first-order integrators, while $\alpha = 1/2$ gives rise to a second-order algorithm. A rather curious aspect of this last remark is that, as mentioned earlier, all the aforementioned algorithms give rise to the *same* DEL, but they differ on the discrete Legendre transform, or the definition of the discrete momenta. Thus, while the coordinate values $\mathbf{q}^i$ coincide in all three algorithms, the momenta $\mathbf{p}^i$ do not, and this is where the order difference between the algorithms comes from.

The midpoint rule gives rise to the same expansion (78) with $\alpha = 1/2$, so its a second-order algorithm.

3.5 Conservation Properties: Discrete Point of View

This section focuses on the geometric properties of the flows generated by variational integrators. As we highlighted before, they correspond to symplectic flows that show an excellent long-term energy behavior along with the exact conservation of the invariants associated to the symmetries of the discrete Lagrangian. These properties along with the existence of a standard methodology to construct high-order methods for Lagrangian systems evolving on general manifolds have contributed to increase the use of VI's in both the scientific and engineering communities.

- **Symmetries and invariants of the dynamics**. A discrete Lagrangian posses a symmetry when it remains invariant under the action of a group on the configuration manifold. Moreover, each symmetry of the discrete Lagrangian leads to a quantity conserved by the dynamics (Marsden and West 2001) according to a discrete version of the celebrated Noether's theorem. This theorem reads as follows,

Discrete version of Noether's theorem. Consider a discrete Lagrangian $L_d(\mathbf{q}^k, \mathbf{q}^{k+1}, h)$ and a one-parameter group of discrete curves $\{\mathbf{q}^{\epsilon,k}\}_{k=0,\ldots,N}$ with $\epsilon > 0$ and $\mathbf{q}^{0,k} = \mathbf{q}^k$, that leaves the discrete Lagrangian invariant in the following sense,

$$L_d\left(\mathbf{q}^{\epsilon,k}, \mathbf{q}^{\epsilon,k+1}, h\right) = L_d\left(\mathbf{q}^k, \mathbf{q}^{k+1}, h\right),$$

for all $\epsilon > 0$ and $k = 0, \ldots, N - 1$. Moreover, consider the infinitesimal symmetry direction,

$$\zeta(\mathbf{q}^k) = \left.\frac{d}{d\epsilon}\mathbf{q}^{\epsilon,k}\right|_{\epsilon=0}.$$

Then

$$I(\mathbf{q}^k, \mathbf{p}^k) = \mathbf{p}^k \cdot \zeta(\mathbf{q}^k) \quad \text{with} \quad \mathbf{p}^k = \frac{\partial L_d}{\partial \mathbf{q}^k}(\mathbf{q}^{k-1}, \mathbf{q}^k, h),$$

is an invariant of the dynamics for all $k = 0, ..., N$.

Proof See, e.g., (Lew et al. 2004) or (Hairer et al. 2006, Chap. VI.6) □

Example Consider the thermoelastic system described in page[7] 20. We know the Lagrangian function is invariant under rigid body translations and rigid body rotations in the physical space and under rigid body translations of the thermal displacements. Therefore, total linear momentum, the total angular momentum and the entropy of each thermoelastic spring are invariants of the dynamics.

We construct a discrete Lagrangian by means of applying the midpoint rule as

$$L_d^m \left(\mathbf{q}^k, \mathbf{q}^{k+1}, \mathbf{\Phi}^k, \mathbf{\Phi}^{k+1}, h\right) = \frac{h}{2}\left(\frac{\mathbf{q}^{k+1}-\mathbf{q}^k}{h}\right) \cdot \mathbf{M}\left(\frac{\mathbf{q}^{k+1}-\mathbf{q}^k}{h}\right) - \\ -h\mathsf{A}\left(\frac{\mathbf{q}^k+\mathbf{q}^{k+1}}{2}, \frac{\mathbf{\Phi}^{k+1}-\mathbf{\Phi}^k}{h}\right). \tag{79}$$

First, we consider the one-parameter group of discrete curves

$$\begin{aligned}\left\{\mathbf{q}_i^{\epsilon,k}(t)\right\}_{k=0,\ldots,N_t} &= \left\{\mathbf{q}_i^k(t) + \epsilon\mathbf{v}\right\}_{k=0,\ldots,N_t} \\ \left\{\mathbf{\Phi}^{k,\epsilon}(t)\right\}_{k=0,\ldots,N_t} &= \left\{\mathbf{\Phi}^k(t) + \epsilon\mathbf{k}\right\}_{k=0,\ldots,N_t},\end{aligned}$$

where $\mathbf{v} \in \mathbb{R}^3$ and $\mathbf{k} \in \mathbb{R}^M$ are constant but otherwise arbitrary vectors. The corresponding infinitesimal directions are given by

$$\boldsymbol{\xi}^k = (\mathbf{v}_N, \mathbf{k}), \qquad k = 0, \ldots, N_t,$$

where $\mathbf{v}_N^k \in \mathbb{R}^d$. Noticing that

$$\begin{aligned}\mathbf{q}^{\epsilon,k+1} - \mathbf{q}^{\epsilon,k} &= \mathbf{q}^{k+1} - \mathbf{q}^k \\ \mathbf{\Phi}^{\epsilon,k+1} - \mathbf{\Phi}^{\epsilon,k} &= \mathbf{\Phi}^{k+1} - \mathbf{\Phi}^k,\end{aligned}$$

and considering that the distance between masses is conserved by rigid body translations in space, i.e.,

$$l_i\left(\frac{\mathbf{q}^{\epsilon,k+1}+\mathbf{q}^{\epsilon,k}}{2}\right) = l_i\left(\frac{\mathbf{q}^{k+1}+\mathbf{q}^k}{2}\right), \qquad i = 1, \ldots, N,$$

[7] In this example N_t denotes the total number of time instants of the discrete trajectory and N is reserved for the number of masses of the thermoelastic system.

it is possible to verify that the discrete Lagrangian remains invariant and thus $\left\{\mathbf{q}^{\epsilon,k}, \Phi^{\epsilon,k}\right\}_{k=0,\dots,N_t}$ is one of its symmetries.

The momentum vector is given by

$$\left(\mathbf{p}^k, \boldsymbol{\eta}^k\right),$$

where

$$\begin{aligned}
\mathbf{p}^k &= \mathbf{M}\left(\frac{\mathbf{q}^{k+1}-\mathbf{q}^k}{h}\right) - \frac{h}{2}\frac{\partial \mathsf{A}}{\partial \mathbf{q}}\left(\frac{\mathbf{q}^{k+1}+\mathbf{q}^k}{2}, \frac{\boldsymbol{\Phi}^{k+1}-\boldsymbol{\Phi}^k}{h}\right)\\
\boldsymbol{\eta}^k &= -\frac{\partial \mathsf{A}}{\partial \boldsymbol{\theta}}\left(\frac{\mathbf{q}^{k+1}+\mathbf{q}^k}{2}, \frac{\boldsymbol{\Phi}^{k+1}-\boldsymbol{\Phi}^k}{h}\right),
\end{aligned}$$

and according to the discrete version of Noether's theorem

$$\mathbf{v}_N \cdot \mathbf{p}^k + \mathbf{k} \cdot \boldsymbol{\eta}^k,$$

is conserved for all $k = 1, \dots, N_t$ and $(\mathbf{v}_N, \mathbf{k}) \in \mathbb{R}^d \times \mathbb{R}^M$. Choosing

$$(\mathbf{v}_N, \mathbf{k}) = ((\boldsymbol{e}_i, \dots, \boldsymbol{e}_i), \mathbf{0}_M), \quad i = 1, 2, 3$$

where $\{\boldsymbol{e}_i\}_{i=1,2,3}$ is a basis in $\mathbb{R}^3$ and $\mathbf{0}_M$ is a M-dimensional vector of zeros, allows to deduce that every component of the total linear momentum is exactly conserved. Moreover, setting

$$(\mathbf{v}_N, \mathbf{k}) = \left(\mathbf{0}_N, \mathbf{k}^j\right),$$

where $\mathbf{k}^j_i = \delta^j_i$ yields to the exact conservation of the entropy of every thermoelastic spring.

Alternatively, we can consider a one-parameter group of discrete curves corresponding to rigid body rotations in the physical space, namely

$$\left\{\mathbf{q}_i^{\epsilon,k}(t)\right\}_{k=0,\dots,N_t} = \left\{\exp\left[\epsilon\widetilde{\boldsymbol{\omega}}\right]\mathbf{q}_i^k(t)\right\}_{k=0,\dots,N_t} \tag{80a}$$

$$\left\{\boldsymbol{\Phi}^{k,\epsilon}(t)\right\}_{k=0,\dots,N_t} = \left\{\boldsymbol{\Phi}^k(t)\right\}_{k=0,\dots,N_t}, \tag{80b}$$

where $\widetilde{\boldsymbol{\omega}}$ is a constant and skew-symmetric but otherwise arbitrary tensor with axial vector $\widehat{\boldsymbol{\omega}} \in \mathbb{R}^3$. The corresponding infinitesimal symmetry direction is given by

$$\zeta(\mathbf{q}^k) = \left\{\left(\left(\widehat{\boldsymbol{\omega}} \times \mathbf{q}_1^k, \dots, \widehat{\boldsymbol{\omega}} \times \mathbf{q}_N^k\right), \mathbf{0}_M\right)\right\}_{k=0,\dots,N_t}.$$

The discrete Lagrangian (79) is invariant under the transformations defined in (80a) and (80b) for all $\epsilon > 0$, since both the discrete version of the kinetic energy and the distance among the masses of the system remain unaffected by rigid body motions in the physical space.

Then, according to the discrete version of Noether's theorem

$$\mathbf{p}^k \cdot \boldsymbol{\xi}(\mathbf{q}^k),$$

is exactly conserved by the dynamics for all $k = 0, \ldots, N_t$. The above expression can be rewritten as

$$\mathbf{A}^k \cdot \widehat{\mathbf{W}} = \sum_{i=1}^{N} \left(\mathbf{q}_i^k \times \left[\mathbf{M} \left(\frac{\mathbf{q}^{k+1} - \mathbf{q}^k}{h} \right) - \frac{\partial \mathsf{A}}{\partial \mathbf{q}_i^k} \left(\mathbf{q}^{k+\frac{1}{2}}, \boldsymbol{\theta}^k \right) \right] \right) \cdot \widehat{\boldsymbol{\omega}},$$

where

$$\widehat{\mathbf{W}} := \underbrace{(\widehat{\boldsymbol{\omega}}, \ldots, \widehat{\boldsymbol{\omega}})}_{N \text{ times}}, \quad \mathbf{q}^{k+\frac{1}{2}} = \frac{\mathbf{q}^{k+1} + \mathbf{q}^k}{2} \quad \text{and} \quad \boldsymbol{\theta}^k = \frac{\boldsymbol{\Phi}^{k+1} - \boldsymbol{\Phi}^k}{h}.$$

Therefore, since $\widehat{\boldsymbol{\omega}}$ is an arbitrary vector in $\mathbb{R}^3$, it is possible to conclude that a discrete version of the total angular momentum, $\mathbf{A}^k$, remains invariant. □

- **Discrete symplecticity**. In Sect. 3.2 we described how the discrete Legendre transform allows to define the discrete Hamiltonian map

$$\begin{array}{rcl} \boldsymbol{\Phi}_h : \; T^*Q & \rightarrow & T^*Q \\ (\mathbf{q}^n, \mathbf{p}^n) & \mapsto & \boldsymbol{\Phi}_h(\mathbf{q}^n, \mathbf{p}^n) = (\mathbf{q}^{n+1}, \mathbf{p}^{n+1}) \end{array},$$

which can be used to construct the position-momentum form of a variational method. In this section we show that $\boldsymbol{\Phi}_h$ also defines a discrete symplectic flow on T^*Q.

Remark Given a particular discrete Lagrangian, to demonstrate that its discrete Hamiltonian map is symplectic, it is enough to verify that $\boldsymbol{\Phi}_h$ fulfils (50), i.e.,

$$\left(\frac{\partial \boldsymbol{\Phi}_h}{\partial \boldsymbol{y}^k} (\boldsymbol{y}^k) \right)^t \mathbf{J}^{-1} \left(\frac{\partial \boldsymbol{\Phi}_h}{\partial \boldsymbol{y}^k} (\boldsymbol{y}^k) \right) = \mathbf{J}^{-1},$$

where $\boldsymbol{y}^k = (\mathbf{q}^k, \mathbf{p}^k) \in T^*Q$. □

However, to prove this in a more general setting, we consider the following result:

Theorem *Any smooth enough and nondegenerate function* $S(\mathbf{q}, \mathbf{Q})$ *generates a symplectic flow* $(\mathbf{q}, \mathbf{p}) \mapsto (\mathbf{P}, \mathbf{Q})$ *if*

$$\mathbf{p} = -\frac{\partial S}{\partial \mathbf{q}}(\mathbf{q}, \mathbf{Q}) \quad \textit{and} \quad \mathbf{P} = \frac{\partial S}{\partial \mathbf{Q}}(\mathbf{q}, \mathbf{Q}). \tag{81}$$

Proof See (Hairer et al. 2006, pp. 196–197) □

The function $S(\mathbf{q}, \mathbf{Q})$ is a particular case of the so-called *generating functions*.

Consider a sequence of points $\{\mathbf{q}^i\}_{i=0,\ldots,N}$ on Q that is the solution of the DEL equations (58) subjected to boundary conditions $\mathbf{q}^0$ and $\mathbf{q}^N$. Then the discrete action sum (55) can be regarded as a function of the initial and final configuration points, i.e.,

$$S_d\left(\mathbf{q}^0, \mathbf{q}^N\right) = \sum_{i=0}^{N-1} L_d\left(\mathbf{q}^i, \mathbf{q}^{i+1}, h\right).$$

Taking into account the discrete Legendre transformations (69a) and (69b), we have that

$$\begin{aligned}\frac{\partial S_d}{\partial \mathbf{q}^0}\left(\mathbf{q}^0, \mathbf{q}^N\right) &= \frac{\partial L_d}{\partial \mathbf{q}^0}\left(\mathbf{q}^0, \mathbf{q}^1, h\right) = -\mathbf{p}^0,\\ \frac{\partial S_d}{\partial \mathbf{q}^N}\left(\mathbf{q}^0, \mathbf{q}^N\right) &= \frac{\partial L_d}{\partial \mathbf{q}^N}\left(\mathbf{q}^{N-1}, \mathbf{q}^N, h\right) = \mathbf{p}^N,\end{aligned}$$

and therefore applying Theorem 3.5, the flow $(\mathbf{q}^0, \mathbf{p}^0) \mapsto (\mathbf{q}^N, \mathbf{p}^N)$ results to be symplectic. This result can be applied to an arbitrary time interval $[t^i, t^{i+1}], i = 0, \ldots, N$, which shows that the discrete flow of any variational integrator is automatically symplectic and that the corresponding generating function is the discrete Lagrangian $L_d(\mathbf{q}^i, \mathbf{q}^{i+1}, h)$.

Example (Hairer et al. 2006, pp. 190). Consider the midpoint discrete Lagrangian,

$$L_d(\mathbf{q}^0, \mathbf{q}^1, h) = \frac{1}{2h}\mathbf{M}(\mathbf{q}^1 - \mathbf{q}^0)\cdot(\mathbf{q}^1 - \mathbf{q}^0) + hU\left(\mathbf{q}^{\frac{1}{2}}\right),$$

where $\mathbf{q}^{\frac{1}{2}} = \frac{1}{2}(\mathbf{q}^1 + \mathbf{q}^0)$. The corresponding variational integrator in position-momentum form is given by

$$\begin{aligned}\mathbf{q}^1 &= \mathbf{q}^0 + h\,\mathbf{M}^{-1}\mathbf{p}^{\frac{1}{2}}\\ \mathbf{p}^1 &= \mathbf{p}^0 - h\,\frac{\partial U}{\partial \mathbf{q}}\left(\mathbf{q}^{\frac{1}{2}}\right),\end{aligned}$$

which can be rewritten as

$$z^1 = z^0 + h\,\mathbf{J}\frac{\partial H}{\partial z}\left(z^{\frac{1}{2}}\right),$$

where $z = (\mathbf{q}, \mathbf{p})$ and $H(z)$ is the Hamiltonian function of the problem. Differentiating the above equation yields

$$\left(\mathbf{I} - \frac{h}{2}\mathbf{J}\frac{\partial^2 H}{\partial z^2}\right)\frac{\partial z^1}{\partial z^0} = \left(\mathbf{I} + \frac{h}{2}\mathbf{J}\frac{\partial^2 H}{\partial z^2}\right),$$

from which it is clear that

$$\left(\frac{\partial z^1}{\partial z^0}\right)^t \mathbf{J}^{-1}\left(\frac{\partial z^1}{\partial z^0}\right) = \mathbf{J}^{-1}.$$

□

- **Long-term energy behavior**. As it has been explained in Sect. 2.5 the flow on the phase space of an autonomous Hamiltonian system is constrained to remain on a constant energy manifold which depends on the initial conditions. Unfortunately, the discrete flows generated by symplectic methods with constant time step cannot conserve exactly the energy of the original Hamiltonian system (Ge and Marsden 1988; Kane et al. 1999). However, in spite of this limitation, they show an excellent long-term behavior with errors in the energy that remain bounded for exponentially long periods of time. See, e.g., (Marsden and West (2001)). In following, we explain the reasons for the superior behavior of symplectic methods.

  In Fig. 15 the numerical flow of the Symplectic-Euler method (obtained from the discrete Lagrangian (54a)) is compared with the numerical flow of the Explicit-Euler method, when both methods are used to simulate the dynamics of the system described by the Lagrangian

$$L(q(t), \dot{q}(t)) = \frac{1}{2}\dot{q}(t)^2 - 5(q(t) - 1)^2,$$

  subjected to the initial conditions $q(0) = 2, \dot{q}(0) = 0$. This system is conservative and thus there should be no loss of energy over time. This figure shows that while

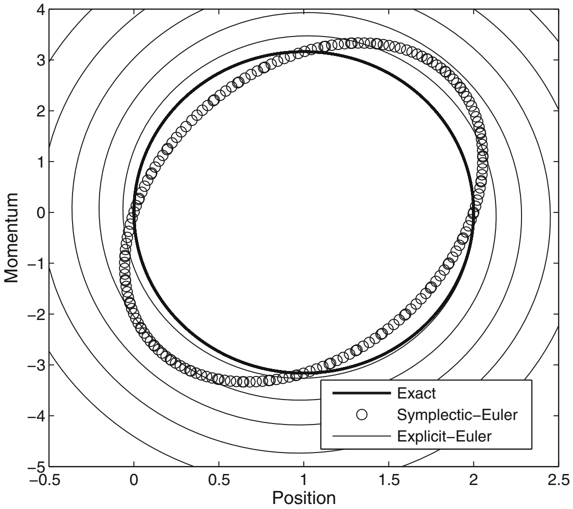


Fig. 15 Comparison between the Symplectic-Euler and the Explicit-Euler methods. Trajectories on the phase space

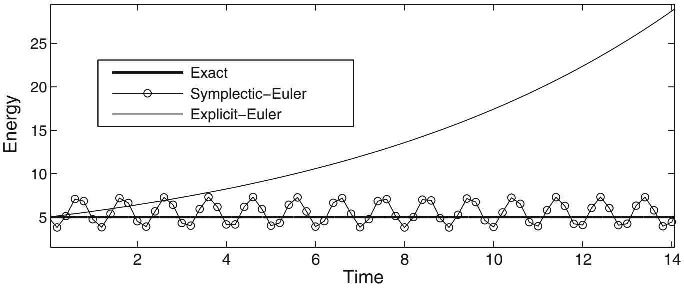


Fig. 16 Comparison between the Symplectic-Euler and the Explicit-Euler methods in terms of the numerically computed energy

the trajectory of the Explicit-Euler method departs progressively from the exact one, the trajectory of the Symplectic-Euler remains bounded and close to the exact trajectory.

In Fig. 16 a plot of the energy

$$H(q(t), p(t)) = \frac{1}{2}m^{-1}p(t)^2 + 5(q(t) - 1)^2,$$

evaluated on the discrete trajectories versus time is shown. The striking aspect of this graph is that while the energy associated with the Explicit-Euler method blows up due to numerical instability, for the Symplectic-Euler method the energy error remains bounded over a long period of time.

Backward error analysis. For a better understanding the above results, we may resort to use the so-called backward error analysis applied to symplectic methods (Marsden and West 2001; Lew et al. 2004; Faltinsen 2000).

Consider a numerical method, here represented by its discrete flow map $\widehat{\mathbf{\Phi}}_h : Q \to Q$, which we use to approximate the flow of the differential equation

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}), \quad \mathbf{y}(0) = \mathbf{y}^0 \in Q.$$

The basic idea of backward error analysis consist constructing *modified differential equation*

$$\widetilde{\mathbf{y}} = \mathbf{f}_h(\widetilde{\mathbf{y}}) = \mathbf{f}(\widetilde{\mathbf{y}}) + h\mathbf{f}_2(\widetilde{\mathbf{y}}) + h^2\mathbf{f}_3(\widetilde{\mathbf{y}}) + ... \tag{82}$$

such that its exact solution trajectory $\widetilde{\mathbf{y}}(t)$ exactly matches the discrete flow of the numerical method, i.e.,

$$\mathbf{y}^k = \widehat{\mathbf{\Phi}}_{kh}(\mathbf{y}^0) = \widetilde{\mathbf{y}}(kh), \quad k = 1, 2, ..., N.$$

Then, the numerical analysis focuses on studying the difference between $\mathbf{f}(\mathbf{y})$ and $\mathbf{f}_h(\mathbf{y})$ in an appropriate norm instead of studying the difference between $\mathbf{y}^k$ and $\widehat{\mathbf{\Phi}}_{kh}(\mathbf{y}^0)$ which is the focus of the more traditional *forward error analysis*. In practice the series (82) diverges and has to be truncated after a finite number of terms. Therefore, this approach allows to interpret the numerical solution of a differential equation as a *higher order approximation* of a modified system. We can now understand why variational integrators are different to standard methods. To this end we first consider the next theorem, which for sake of simplicity is restricted to Hamiltonian functions taking arguments in $\mathbb{R}^{2d}$.

Theorem *The modified equation of a symplectic method* $\mathbf{\Phi}_h$ *applied to a Hamiltonian system with a smooth Hamiltonian* $H : \mathbb{R}^{2d} \to \mathbb{R}$ *is also Hamiltonian. It means that there exist smooth functions* $H_j : \mathbb{R}^{2d} \to \mathbb{R}$ *for* $j = 2, 3, ...,$ *such that*

$$\dot{\mathbf{y}} = \mathbf{J}\left(\frac{\partial H}{\partial \mathbf{y}}(\mathbf{y}) + h\frac{\partial H_2}{\partial \mathbf{y}}(\mathbf{y}) + h^2\frac{\partial H_3}{\partial \mathbf{y}}(\mathbf{y}) + \cdots\right).$$

Proof See, e.g., (Hairer et al. 2006). □

In other words, $\mathbf{\Phi}_h$ is a higher order approximation to the flow of the dynamical system defined by a *shadow Hamiltonian*

$$\widetilde{H}_h(\mathbf{y}) = H(\mathbf{y}) + \sum_{i=2}^{N} h^{i-1} H_i(\mathbf{y}),$$

which remains at least $\mathcal{O}(h)$ close to H.

Since every Lagrangian system admits a Hamiltonian representation, the modified differential equation of every variational integrator is Hamiltonian. This means that the discrete trajectory has all of the properties of a conservative mechanical system, such as energy conservation. This property explains the shape of the closed trajectory described by the Symplectic-Euler in Fig. 15. This also explains why the energy plots for variational integrators contain a typical oscillation about a value close to the true energy. See Fig. 16. The modified energy level set will be close to the true energy level set everywhere, but it will typically be inside it at some locations and outside it at others.

Finally, we mention a few words regarding to how much the discrete trajectory moves away from both the exact energy manifold and the constant energy manifold defined by the shadow Hamiltonian. If we apply an order p numerical method $\mathbf{\Phi}_h$ with step size h to approximate the flow of a Hamiltonian system with analytic $H : D \subset \mathbb{R}^{2d} \to \mathbb{R}$, and if the numerical solution stays in the compact set $K \subset D$, then there exist h_0 and $N = N(h)$ such that

$$\widetilde{H}(\mathbf{y}^k) = \widetilde{H}(\mathbf{y}^0) + \mathcal{O}\left(e^{h_0/2h}\right),$$
$$H(\mathbf{y}^k) = H(\mathbf{y}^0) + \mathcal{O}(h^p),$$

over exponentially long time intervals $kh \leq e^{h_0/2h}$. See (Hairer et al. 2006, Chap. IX) for details.

Therefore, it is possible to see that the discrete trajectory remains exponentially close to the constant energy manifold defined by the shadow Hamiltonian for exponentially longs periods of time. Moreover, it also remains $\mathcal{O}(h^p)$ close to the manifold of exact energy.

4 Final Examples

In this section, we formulate variational time integrators for two problems of practical interest in science and engineering. First, we develop a second-order method based on the trapezoidal rule for a free-flying body that is able to undergo arbitrarily large rotations and displacements in space. This problem has been extensively studied from both theoretical and numerical aspects (see, e.g., Meyer et al. 2009; Bauchau and Bottasso 1999; Chaturvedi et al. 2011; Lee et al. 2007; Marsden and Ratiu 1999; Simo and Wong 1991) among others) since its configuration space corresponds to a nonlinear differentiable manifold rather than a linear space. The second example corresponds to the formulation of an explicit, second-order accurate varitational integrator for finite element discretizations of geometrically exact rods. We only consider linear finite elements in space since a more general formulation can be consulted in (Mata 2015). In both examples the time interval of interest $[0, T]$ is partitioned into $N > 1$ subintervals with constant time step $\Delta t = T/N$, and we set $t^k = kT/N$, $k = 0, 1, ..., N$.

4.1 Rotating Rigid Body

Consider an inertial reference frame $\{\boldsymbol{e}_i\}_{i=1,2,3}$ in the three-dimensional space and a rigid body $\mathcal{B}$ with mass density $\rho > 0$ which has rigidly attached to its center of mass an orthogonal reference frame $\{\boldsymbol{t}_i\}_{i=1,2,3}$. See Fig. 17. The orientation of the body-fixed frame with respect to the inertial frame is specified by means of a rotation tensor $\boldsymbol{\Lambda}$ according to

$$\boldsymbol{t}_i = \boldsymbol{\Lambda}\boldsymbol{e}_i, \qquad i = 1, 2, 3.$$

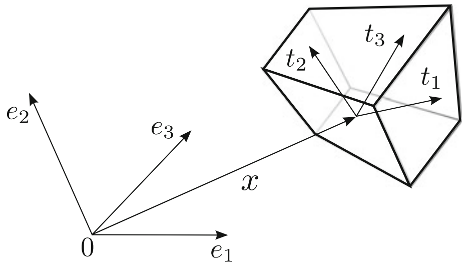


Fig. 17 Free-flying rigid body. $\{\boldsymbol{e}_i\}_{i=1,2,3}$ is a inertial reference frame and $\{\boldsymbol{t}_i\}_{i=1,2,3}$ an orthogonal body-fixed reference frame

The position vector of a material point of $\mathcal{B}$ is given by

$$\boldsymbol{y} = \boldsymbol{x} + \sum_{i=1}^{3} \xi_i \boldsymbol{t}_i = \boldsymbol{x} + \sum_{i=1}^{3} \xi_i \boldsymbol{\Lambda} \boldsymbol{e}_i,$$

where $\boldsymbol{\xi} = (\xi^1, \xi^2, \xi^3)$ is a set of coordinates with respect to $\{\boldsymbol{t}_i\}_{i=1,2,3}$ and $\boldsymbol{x} = x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 + x^3\boldsymbol{e}_3$ is the position vector of the center of mass of the body. The spatial position of the body is specified by the pair $(\boldsymbol{x}, \boldsymbol{\Lambda})$ which is composed of a position vector plus a rotation tensor measuring the deviation of body-fixed frame with respect to the inertial reference frame.

Note that although $\boldsymbol{x}$ belongs to $\mathbb{R}^3$ which is a linear space, $\boldsymbol{\Lambda}$ is an element of the noncommutative (Lie) group of proper rotations

$$SO(3) = \left\{ \boldsymbol{\sigma} \in \mathbb{R}^{3\times 3} \ \mid \ \boldsymbol{\sigma}^{-1} = \boldsymbol{\sigma}^t \quad \text{and} \quad \det[\boldsymbol{\sigma}] = 1 \right\},$$

which is a nonlinear manifold. A brief introduction to finite rotations is given in Sect. B. Then, the configuration the manifold is $SE(3) = \mathbb{R}^3 \times SO(3)$.

A motion of the body can be described by means of the time-dependent curve

$$\boldsymbol{\Phi} = (\boldsymbol{x}, \boldsymbol{\Lambda}) : [0, T] \to SE(3), \tag{83}$$

with velocity given by

$$\dot{\boldsymbol{\Phi}} = (\dot{\boldsymbol{x}}, \dot{\boldsymbol{\Lambda}}) \in T_{(\boldsymbol{x},\boldsymbol{\Lambda})} SE(3) = \mathbb{R}^3 \times T_{\boldsymbol{\Lambda}} SO(3).$$

For free-flying bodies the Lagrangian function $L : TSE(3) \to \mathbb{R}$ is equal to the kinetic energy

$$L(\boldsymbol{\Phi}, \dot{\boldsymbol{\Phi}}) = \frac{1}{2} \left(m\, \dot{\boldsymbol{x}} \cdot \dot{\boldsymbol{x}} + \operatorname{tr}\left[\widetilde{\boldsymbol{\Omega}} \mathbf{J}_d \widetilde{\boldsymbol{\Omega}}^t \right] \right), \tag{84}$$

where $m = \int_{\mathcal{B}} \rho \, dv$ is the total mass of the body,

$$\widetilde{\boldsymbol{\Omega}} = \boldsymbol{\Lambda}^t \dot{\boldsymbol{\Lambda}},$$

is the angular velocity tensor expressed in the body-fixed frame which belongs to the (linear) space of skew-symmetric tensors $so(3)$, and

$$\mathbf{J}_d = \int_{\mathcal{B}} \rho\, \boldsymbol{\xi} \otimes \boldsymbol{\xi} dv,$$

is a nonstandard moment of inertia tensor which is related to the standard symmetric moment of inertia tensor, $\mathbf{J}$, by

$$\mathbf{J} = \int_{\mathcal{B}} \rho\, \widetilde{\boldsymbol{\xi}}^t \widetilde{\boldsymbol{\xi}}\, dv = \mathrm{tr}\,[\mathbf{J}_d]\,\mathbf{I} - \mathbf{J}_d,$$

where $\widetilde{\boldsymbol{\xi}} = \mathrm{skew}[\boldsymbol{\xi}] \in so(3)$ is the skew-symmetric tensor obtained from $\boldsymbol{\xi} \in \mathbb{R}^3$. See Sect. B.3. More details about this relation can be found in (Lee et al. 2007).

The application of Hamilton's principle requires computing

$$\Big\langle \delta S[\boldsymbol{\Phi}(t)], \delta\boldsymbol{\Phi}(t) \Big\rangle = \frac{dL}{d\epsilon}\left(\boldsymbol{x}_\epsilon(t), \boldsymbol{\Lambda}_\epsilon(t), \dot{\boldsymbol{x}}_\epsilon(t), \dot{\boldsymbol{\Lambda}}_\epsilon(t)\right) = 0, \tag{85}$$

where $\delta\boldsymbol{\Phi}(t) = (\delta\boldsymbol{x}(t), \delta\boldsymbol{\Lambda}(t))$ represents a variation over an arbitrary element $\boldsymbol{\Phi}(t)$ belonging to the set $\mathcal{C}$ composed by all the smooth enough trajectories of the form (83) that leaves the endpoints of the trajectory fixed. This is by no means a trivial task owing to the nonlinear nature of $SE(3)$. On one hand, we have that

$$\begin{aligned} \delta\boldsymbol{x}(t) &= \frac{d}{d\epsilon}\left(\boldsymbol{x}(t) + \epsilon\boldsymbol{u}(t)\right)\Big|_{\epsilon=0} = \boldsymbol{u}(t), \\ \delta\boldsymbol{\Lambda}(t) &= \frac{d}{d\epsilon}\left(\exp\left[\epsilon\widetilde{\boldsymbol{\Theta}}(t)\right]\boldsymbol{\Lambda}(t)\right)\Big|_{\epsilon=0} = \widetilde{\boldsymbol{\Theta}}(t)\boldsymbol{\Lambda}(t), \end{aligned}$$

where

$$\boldsymbol{u} : [0, T] \to \mathbb{R}^3 \quad \text{and} \quad \widetilde{\boldsymbol{\Theta}} : [0, T] \to so(3).$$

Therefore, a variation over $\boldsymbol{\Phi}(\cdot)$ is given by

$$\delta\boldsymbol{\Phi} = \left(\boldsymbol{u}, \widetilde{\boldsymbol{\Theta}}\boldsymbol{\Lambda}\right) \in T_{\boldsymbol{\Phi}}\mathcal{C}.$$

The above results allow to see that

$$\delta\dot{\boldsymbol{x}} = \dot{\boldsymbol{u}} \qquad \text{and} \qquad \delta\widetilde{\boldsymbol{\Omega}} = \boldsymbol{\Lambda}^t\dot{\widetilde{\boldsymbol{\Theta}}}\boldsymbol{\Lambda}.$$

Replacing the above results in (85) yields

$$\Big\langle \delta S[\boldsymbol{\Phi}], \delta\boldsymbol{\Phi} \Big\rangle = \frac{1}{2}\int_0^T \mathrm{tr}\left[\delta\widetilde{\boldsymbol{\Theta}}\left(\frac{d}{dt}(\mathbf{j}_d\widetilde{\boldsymbol{\omega}} + \widetilde{\boldsymbol{\omega}}\mathbf{j}_d)\right)\right] dt - \int_0^T m\,\ddot{\boldsymbol{x}} \cdot \boldsymbol{u}\, dt = 0,$$

where $\widetilde{\boldsymbol{\omega}} = \boldsymbol{\Lambda}\widetilde{\boldsymbol{\Omega}}\boldsymbol{\Lambda}^t$ and $\mathbf{j}_d = \boldsymbol{\Lambda}\mathbf{J}_d\boldsymbol{\Lambda}^t$ are the spatial forms of the angular velocity tensor and nonstandard inertia tensor, respectively. Noting that since $(\boldsymbol{u}(t), \widetilde{\boldsymbol{\Theta}}(t))$ are arbitrary for all $t \in [0, T]$ and that

$$\text{skew}\big[\mathbf{j}\,\boldsymbol{\omega}\big] = \mathbf{j}_d\,\widetilde{\boldsymbol{\omega}} + \widetilde{\boldsymbol{\omega}}\,\mathbf{j}_d, \qquad (\widetilde{\boldsymbol{\omega}} = \text{skew}[\boldsymbol{\omega}]),$$

the following system of Euler–Lagrange equations is obtained

$$m\ddot{\boldsymbol{x}} = 0 \quad \text{and} \quad \frac{d}{dt}\left(\text{skew}\,[\mathbf{j}\,\boldsymbol{\omega}]\right) = 0,$$

where $\boldsymbol{\omega} \in \mathbb{R}^3$ is the axial vector of $\widetilde{\boldsymbol{\omega}}$ and $\mathbf{j} = \boldsymbol{\Lambda}\mathbf{J}\boldsymbol{\Lambda}^t$. Basically, both equations are alternative statements for the conservation of the total linear momentum and total angular momentum.

4.1.1 Legendre Transforms

The momentum vectors are computed with the help of the Legendre transformation as

$$\mathbf{p} = m\dot{\boldsymbol{x}}, \qquad \text{and} \qquad \boldsymbol{\pi} = \mathbf{j}\boldsymbol{\omega},$$

which allows to define the Hamiltonian function, $H : T^*SE(3) \to \mathbb{R}$, as

$$H(\mathbf{p}, \boldsymbol{\pi}) = \frac{1}{2}\left(m^{-1}\mathbf{p}\cdot\mathbf{p} + \mathbf{j}^{-1}\boldsymbol{\pi}\cdot\boldsymbol{\pi}\right),$$

and therefore, the E–L equations may be rewritten as

$$\dot{\mathbf{p}} = 0 \qquad \text{and} \qquad \dot{\tilde{\boldsymbol{\pi}}} = 0, \tag{86}$$

respectively.

4.1.2 Discretization

The procedure to construct a variational integrator for this problem follows some ideas presented in (Lee et al. 2007) for the full-body problem. We denote by $\boldsymbol{x}^k \approx \boldsymbol{x}(t^k)$ and $\boldsymbol{\Lambda}^k \approx \boldsymbol{\Lambda}(t^k)$, $k = 1, ..., N$ and we assume that the velocities are approximated by

$$\dot{\boldsymbol{x}}(\tau) \approx \frac{\boldsymbol{x}^1 - \boldsymbol{x}^0}{h}, \tag{87a}$$

$$\widetilde{\boldsymbol{\Omega}}(\tau) \approx \boldsymbol{\Lambda}^{0t}\frac{\boldsymbol{\Lambda}^1 - \boldsymbol{\Lambda}^0}{h}, \qquad \tau \in [t^0, t^1], \tag{87b}$$

and therefore,

$$\operatorname{tr}\left[\widetilde{\boldsymbol{\Omega}}\mathbf{J}_d\widetilde{\boldsymbol{\Omega}}^t\right] \approx \operatorname{tr}\left[\boldsymbol{\Lambda}^{0t}\frac{\boldsymbol{\Lambda}^1-\boldsymbol{\Lambda}^0}{h}\mathbf{J}_d\left(\boldsymbol{\Lambda}^{0t}\frac{\boldsymbol{\Lambda}^1-\boldsymbol{\Lambda}^0}{h}\right)^t\right], \tag{88}$$

which, taking into account that for any two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, $\operatorname{tr}[\mathbf{AB}] = \operatorname{tr}[\mathbf{BA}] = \operatorname{tr}[\mathbf{A}^t\mathbf{B}^t]$, yields

$$\operatorname{tr}\left[\widetilde{\boldsymbol{\Omega}}\mathbf{J}_d\widetilde{\boldsymbol{\Omega}}^t\right] \approx \frac{2}{h^2}\operatorname{tr}\left[\left(\mathbf{I}-\boldsymbol{\Lambda}^{0t}\boldsymbol{\Lambda}^1\right)\mathbf{J}_d\right].$$

Then, the discrete Lagrangian based on the trapezoidal rule is given by

$$L_d(\boldsymbol{x}^0,\boldsymbol{x}^1,\boldsymbol{\Lambda}^0,\boldsymbol{\Lambda}^1) = \frac{1}{2h}m(\boldsymbol{x}^1-\boldsymbol{x}^0)\cdot(\boldsymbol{x}^1-\boldsymbol{x}^0) + \frac{1}{h}\operatorname{tr}\left[\left(\mathbf{I}-\boldsymbol{\Lambda}^{0t}\boldsymbol{\Lambda}^1\right)\mathbf{J}_d\right]. \tag{89}$$

Remark If the midpoint rule is preferred, the following approximation has to be used for the angular velocity tensor

$$\widetilde{\boldsymbol{\Omega}} \approx \frac{\boldsymbol{\Lambda}^{0t}+\boldsymbol{\Lambda}^{1t}}{2}\frac{\boldsymbol{\Lambda}^1-\boldsymbol{\Lambda}^0}{h}.$$

Moreover, denoting by

$$\begin{aligned}\mathbf{A} &= (\boldsymbol{\Lambda}^{0t}+\boldsymbol{\Lambda}^{1t})(\boldsymbol{\Lambda}^1-\boldsymbol{\Lambda}^0)\mathbf{J}_d\\ \mathbf{B} &= (\boldsymbol{\Lambda}^{1t}-\boldsymbol{\Lambda}^{0t})(\boldsymbol{\Lambda}^0+\boldsymbol{\Lambda}^1)\end{aligned}$$

and applying the fact that $\operatorname{tr}[\mathbf{AB}] = \operatorname{tr}\left[\mathbf{A}^t\mathbf{B}^t\right]$, it is possible to obtain

$$\operatorname{tr}\left[\widetilde{\boldsymbol{\Omega}}\mathbf{J}_d\widetilde{\boldsymbol{\Omega}}^t\right] \approx \frac{1}{2h^2}\operatorname{tr}\left[\mathbf{J}_d\left(\mathbf{I}-\boldsymbol{\Lambda}^{1t}\boldsymbol{\Lambda}^0\boldsymbol{\Lambda}^{1t}\boldsymbol{\Lambda}^0\right)\right],$$

which contains a higher power of $(\boldsymbol{\Lambda}^{1t}\boldsymbol{\Lambda}^0)$ ■

From (89) it is possible to define the discrete translational momenta as

$$\mathbf{p}^0 = -D_{\boldsymbol{x}^0}L_d = \frac{m}{h}(\boldsymbol{x}^1-\boldsymbol{x}^0), \tag{90a}$$

$$\mathbf{p}^1 = D_{\boldsymbol{x}^1}L_d = \mathbf{p}^0. \tag{90b}$$

Determining the momenta associated to the rotational part of the motion is a little bit more involved. On one hand, considering that $\delta\boldsymbol{\Lambda}^0 = \widetilde{\boldsymbol{\Theta}}^0\boldsymbol{\Lambda}^0$, we have that the following relation holds

$$\operatorname{tr}\left[-\left(D_{\boldsymbol{\Lambda}^0}L_d\right)\delta\boldsymbol{\Lambda}^0\right] = \operatorname{tr}\left[-\widetilde{\boldsymbol{\Theta}}^0\boldsymbol{\Lambda}^0 D_{\boldsymbol{\Lambda}^0}L_d\right] = \operatorname{tr}\left[\widetilde{\boldsymbol{\Theta}}^0\left(\frac{1}{h}\boldsymbol{\Lambda}^1\mathbf{J}_d\boldsymbol{\Lambda}^{0t}\right)\right] = 0.$$

On the other hand, we know that $\widetilde{\mathbf{\Pi}}^0$, the momentum conjugated to the rotation $\mathbf{\Lambda}^0$, belongs to the linear space $so(3)^*$ which is the dual space of $so(3)$. Moreover, $so(3)^*$ is also composed by skew-symmetric tensors. See, e.g., (Wendlandt and Marsden 1997; Marsden and Ratiu 1999).

Therefore, since $\widetilde{\mathbf{\Theta}}^0$ is skew-symmetric and

$$\mathrm{tr}\left[\widetilde{\mathbf{\Theta}}^0\mathbf{B}\right] = \frac{1}{2}\mathrm{tr}\left[\widetilde{\mathbf{\Theta}}^0(\mathbf{B}-\mathbf{B}^t)\right],$$

for any 3 × 3 matrix $\mathbf{B}$, we have that the above equation can be rewritten as

$$\frac{1}{h}\mathrm{tr}\left[\widetilde{\mathbf{\Theta}}^0\left(\mathbf{\Lambda}^1\mathbf{J}_d\mathbf{\Lambda}^{0t} - \mathbf{\Lambda}^0\mathbf{J}_d\mathbf{\Lambda}^{1t}\right)\right] = \frac{1}{h}\mathrm{tr}\left[\widetilde{\mathbf{\Theta}}^0\widetilde{\mathbf{\Pi}}^0\right],$$

which allows to identify

$$\widetilde{\mathbf{\Pi}}^0 = \frac{1}{h}\left(\mathbf{\Lambda}^{in}\mathbf{j}_d^0 - \mathbf{j}_d^0\mathbf{\Lambda}^{in(t)}\right), \tag{90c}$$

with the momentum conjugated to $\mathbf{\Lambda}^0$ after defining $\mathbf{j}_d^0 = \mathbf{\Lambda}^0\mathbf{J}_d\mathbf{\Lambda}^{0t}$ and $\mathbf{\Lambda}^{in} = \mathbf{\Lambda}^1\mathbf{\Lambda}^{0t}$. On the other hand, following an analogue procedure allows to compute the skew-symmetric momentum associated to $\mathbf{\Lambda}^1$ as

$$\widetilde{\mathbf{\Pi}}^1 = \widetilde{\mathbf{\Pi}}^0. \tag{90d}$$

Note that (90b) and (90d) are the DEL equations. They also are discrete counterparts of the conservation laws $(86)^{1,2}$.

4.1.3 Solution Procedure

Since the translational and rotational momenta are exactly conserved by the algorithm, we only need to determine $(\boldsymbol{x}^1, \mathbf{\Lambda}^1)$ from $(\boldsymbol{x}^0, \mathbf{\Lambda}^0, \mathbf{p}^0, \mathbf{\Pi}^0)$. The procedure is as follows.

- The position in space is updated using (90a) as

$$\boldsymbol{x}^1 = \boldsymbol{x}^0 + \frac{h}{m}\mathbf{p}^0.$$

- To determine $\mathbf{\Lambda}^1$ we parametrize the incremental rotation tensor $\mathbf{\Lambda}^{in}$ in terms of an incremental rotation vector $\boldsymbol{\theta} \in \mathbb{R}^3$ as

$$\mathbf{\Lambda}^{in} = \exp[\widetilde{\boldsymbol{\theta}}],$$

where $\widetilde{\boldsymbol{\theta}} = \text{skew}[\boldsymbol{\theta}]$ is the skew-symmetric tensor obtained from $\boldsymbol{\theta}$ and $\exp[\bullet] : so(3) \mapsto SO(3)$ is the exponential map (see Sect. B.3). Then, replacing in (90c) yields to the following nonlinear system of equations

$$\left(\frac{\sin\theta}{\theta}\right)\mathbf{j}_d^0\,\boldsymbol{\theta} + \left(\frac{1-\cos\theta}{\theta^2}\right)\boldsymbol{\theta}\times\mathbf{j}_d^0\,\boldsymbol{\theta} = \boldsymbol{\Pi}^0, \tag{91}$$

where $\theta = \|\boldsymbol{\theta}\|$. This system is solved with the help of the Newton–Raphson scheme. Having obtained $\boldsymbol{\theta}$, we update $\boldsymbol{\Lambda}^1 = \boldsymbol{\Lambda}^{in}\boldsymbol{\Lambda}^0$.

4.2 A Model for Geometrically Exact Rods

In this example, we take advantage of some of the previous results to build an explicit time integrator for finite element discretizations of geometrically exact rods made of an isotropic, homogeneous and hyperelastic material.

4.2.1 Continuum Model

First, we briefly review some basic results of the continuum model. The reference configuration corresponds to a straight rod of length L and constant cross-section $\mathcal{A} \subset \mathbb{R}^2$. The position vector of a material point in this configuration is

$$\mathbf{X}(s, \xi_1, \xi_2) = s\mathbf{E}_1 + \xi_2\mathbf{E}_2 + \xi_3\mathbf{E}_3,$$

where $\{\mathbf{E}_i\}_{i=1,..,3}$ is an (orthogonal) inertial reference frame, $s \in [0, L]$ is an arch-length coordinate and $(\xi_2, \xi_3) \subset \mathcal{A}$ are coordinates on the cross section. The geometric place of points of the form $\mathbf{X}(s, 0, 0)$ defines a reference curve $\boldsymbol{\varphi}_0$. The current configuration of the rod is characterized by the fields

$$\boldsymbol{\Phi} = (\boldsymbol{\varphi}, \boldsymbol{\Lambda}) : [0, L] \to \mathbb{R}^3 \times SO(3), \tag{92}$$

where $\boldsymbol{\varphi}$ is obtained by adding a displacement field onto $\boldsymbol{\varphi}_0$ and $\boldsymbol{\Lambda}$ defines the orientation of the reference frame

$$\boldsymbol{t}_i(s) = \boldsymbol{\Lambda}(s)\mathbf{E}_i, \qquad s \in [0, L], \quad i = 1, 2, 3,$$

which rigidly attached to the cross-section at $s \in [0, L]$ and oriented along the principal axis of inertia by convenience. The position vector of a material point in the current configuration is given by

$$\mathbf{x}(s, \xi_2, \xi_3) = \boldsymbol{\varphi}(s) + \xi_2\mathbf{t}_2(s) + \xi_3\mathbf{t}_3(s).$$

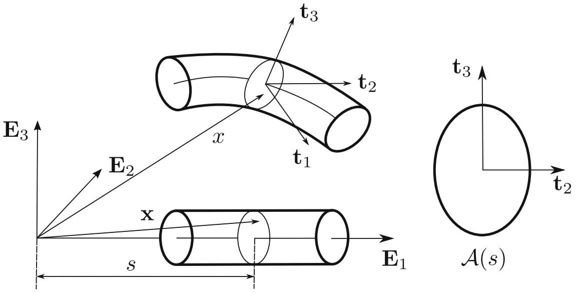


Fig. 18 Reference and current configurations of the elastic rod

See e.g., (Simo 1985; Simo and Vu-Quoc 1986, 1988; Kapania and Li 2003) and Fig. 18.

Then, the configuration manifold, $\mathcal{Q}$, is the set of all the smooth-enough fields of the form (92) subjected to the prescribed boundary conditions $\mathbf{\Phi}(0) = \mathbf{\Phi}_0$ and $\mathbf{\Phi}(L) = \mathbf{\Phi}_L$ and to the restriction $\frac{\partial \varphi}{\partial s} \cdot \mathbf{t}_1 > 0$ (Simo et al. 1995).

Given a motion $\mathbf{\Phi}(\cdot) : [0, T] \to \mathcal{Q}$, the corresponding velocity is obtained as $\dot{\mathbf{\Phi}} = (\dot{\boldsymbol{\varphi}}, \dot{\mathbf{\Lambda}})$ which yields to the following expression for the velocity of a material point:

$$\dot{\mathbf{x}} = \dot{\boldsymbol{\varphi}} + \mathbf{\Lambda}\widetilde{\mathbf{V}}\mathbf{Z}, \tag{93}$$

where $\mathbf{Z} = \xi_2\mathbf{E}_2 + \xi_3\mathbf{E}_3$ and $\widetilde{\mathbf{V}} = \mathbf{\Lambda}^{\mathrm{t}}\dot{\mathbf{\Lambda}}$, is the material form of the angular velocity tensor.

Moreover, consider a stored energy function per unit of reference length $\psi(\mathbf{\Gamma}, \mathbf{\Omega})$ such that the material form of the cross-sectional stress resultants and stress couples are given by

$$\mathbf{n}^{\mathrm{m}} = \frac{\partial \psi}{\partial \mathbf{\Gamma}}(\mathbf{\Gamma}, \mathbf{\Omega}) = \mathbf{C}_\Gamma \mathbf{\Gamma}, \quad \text{and} \quad \mathbf{m}^{\mathrm{m}} = \frac{\partial \psi}{\partial \mathbf{\Omega}}(\mathbf{\Gamma}, \mathbf{\Omega}) = \mathbf{C}_\Omega \mathbf{\Omega}, \tag{94}$$

where

$$\mathbf{\Gamma} = \mathbf{\Lambda}^t(\frac{\partial \boldsymbol{\varphi}}{\partial s} - \boldsymbol{t}_1)$$

is the (translational) strain vector, $\mathbf{\Omega}$ is the axial vector of the curvature tensor

$$\widetilde{\mathbf{\Omega}} = \mathbf{\Lambda}^{\mathrm{t}}\frac{\partial \mathbf{\Lambda}}{\partial s}$$

and the constitutive tensors $\mathbf{C}_\Gamma$ and $\mathbf{C}_\Omega$ are given by

$$\mathbf{C}_\Gamma = \begin{bmatrix} E\mathcal{A} & 0 & \\ 0 & G\mathcal{A} & 0 \\ 0 & 0 & G\mathcal{A} \end{bmatrix} \quad \text{and} \quad \mathbf{C}_\Omega = \begin{bmatrix} GJ & 0 & 0 \\ 0 & EI_2 & 0 \\ 0 & 0 & EI_3 \end{bmatrix},$$

where E, G, J, I_2, and I_3 are an elastic modulus, a shear modulus, a torsional stiffness, and two flexural stiffnesses, respectively. We also define the spatial forms of the stress resultants and couples as

$$\mathbf{n} = \boldsymbol{\Lambda}\mathbf{n}^{\mathrm{m}} \quad \text{and} \quad \mathbf{m} = \boldsymbol{\Lambda}\mathbf{m}^{\mathrm{m}},$$

respectively.

4.2.2 Hamilton's Principle

The Lagrangian function $\mathsf{L} : T\mathcal{Q} \to \mathbb{R}$ is constructed as the kinetic minus the potential energy of the system, i.e.,

$$\mathsf{L}(\boldsymbol{\Phi}, \dot{\boldsymbol{\Phi}}) = \mathsf{K}(\dot{\boldsymbol{\Phi}}, \dot{\boldsymbol{\Phi}}) - \mathsf{U}(\boldsymbol{\Phi}), \tag{95a}$$

where

$$\mathsf{K}(\boldsymbol{\Phi}, \dot{\boldsymbol{\Phi}}) = \frac{1}{2}\int_0^L \left(\mathcal{A}_\rho \dot{\boldsymbol{\varphi}} \cdot \dot{\boldsymbol{\varphi}} - \mathrm{Tr}[\widetilde{\mathbf{V}}\mathbf{E}_\rho \widetilde{\mathbf{V}}]\right) ds, \tag{95b}$$

$$\mathsf{U}(\boldsymbol{\Phi}) = \int_0^L \psi(\boldsymbol{\Gamma}, \boldsymbol{\Omega})\, ds. \tag{95c}$$

In (95b) the mass density per unit of reference length is given by

$$\mathcal{A}_\rho = \int_{\mathcal{A}} \rho_0 \, dA$$

and

$$\mathbf{E}_\rho = \int_{\mathcal{A}} \rho_0 \mathbf{Z} \otimes \mathbf{Z} \, dA,$$

a cross-sectional nonstandard inertia tensor. Compare with (84).

Consider the set $\mathcal{C}$ composed by all the smooth enough motions $\boldsymbol{\Phi} : [0, T] \to \mathcal{Q}$. Hamilton's principle states that the trajectory followed by the system is a stationary point of the action under all variations in $\mathcal{C}$ that leaves fixed the end points $\boldsymbol{\Phi}(0) = \boldsymbol{\Phi}(T) = \mathbf{0}$. This principle yields to the following Euler–Lagrange equations

$$\mathcal{A}_\rho \frac{d^2\boldsymbol{\varphi}}{dt^2} = \frac{\partial \mathbf{n}}{\partial s} + \mathbf{N}, \tag{96a}$$

$$\frac{d}{dt}\left(\mathrm{skew}[\mathbf{i}_\rho \mathbf{v}]\right) = \frac{\partial \widetilde{\mathbf{m}}}{\partial s} + \frac{\partial \widetilde{\boldsymbol{\varphi}}}{\partial s}\widetilde{\mathbf{n}} - \widetilde{\mathbf{n}}\frac{\partial \widetilde{\boldsymbol{\varphi}}}{\partial s} + \widetilde{\mathbf{M}}, \tag{96b}$$

which have to be supplemented with adequate initial conditions. In the above equations $\mathbf{i}_\rho$ is the spatial form of the inertial tensor, $\mathbf{N}$ a vector of the external forces and $\widetilde{\mathbf{M}}$ a skew-symmetric tensor of external moments.

4.2.3 Hamiltonian Framework

The Momentum densities

$$\mathbf{p} = \mathcal{A}_\rho \dot{\boldsymbol{\varphi}} \qquad \text{and} \qquad \widetilde{\boldsymbol{\pi}} = \text{skew}\left[\mathbf{i}_\rho \mathbf{v}\right], \tag{97}$$

are introduced by means of the Legendre transforms. Rewriting the kinetic energy in terms of $(\mathbf{p}, \boldsymbol{\pi})$ it is possible to define the Hamiltonian function $\mathsf{H} : \mathsf{T}^*\mathcal{Q} \to \mathbb{R}$ as

$$\mathsf{H}(\mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\Phi}) = \mathsf{K}(\mathbf{p}, \boldsymbol{\pi}) + \mathsf{U}(\boldsymbol{\Phi}) = \frac{1}{2}\int_0^L \left(\mathcal{A}_\rho^{-1}\mathbf{p}\cdot\mathbf{p} + \mathbf{i}_\rho^{-1}\boldsymbol{\pi}\cdot\boldsymbol{\pi}\right) ds + \mathsf{U}(\boldsymbol{\Phi}),$$

from which it is possible to obtain the balance equations in Hamiltonian form as

$$\begin{aligned}
\dot{\boldsymbol{\varphi}} &= \mathcal{A}_\rho^{-1}\mathbf{p},\\
\dot{\boldsymbol{\Lambda}} &= \text{skew}[\mathbf{i}_\rho^{-1}\boldsymbol{\pi}]\boldsymbol{\Lambda},\\
\dot{\mathbf{p}} &= \frac{\partial \mathbf{n}}{\partial s} + \mathbf{N},\\
\dot{\widetilde{\boldsymbol{\pi}}} &= \frac{\partial \widetilde{\mathbf{m}}}{\partial s} + \frac{\partial \widetilde{\boldsymbol{\varphi}}}{\partial s}\widetilde{\mathbf{n}} - \widetilde{\mathbf{n}}\frac{\partial \widetilde{\boldsymbol{\varphi}}}{\partial s} + \tilde{\mathbf{M}}.
\end{aligned}$$

Regarding to the invariants of the dynamics, we note that the Lagrangian is invariant under translations in time and under rigid body translations and rotations in space. These properties yields to the conservation of energy along with conservation of the total linear momentum, $\mathbf{L}_\mathrm{t}$, and total angular momentum, $\mathbf{J}_\mathrm{t}$, which are given by

$$\mathbf{L}_\mathrm{t} = \int_0^L \mathbf{p}\, ds \qquad \text{and} \qquad \mathbf{J}_\mathrm{t} = \int_0^L (\boldsymbol{\varphi} \times \mathbf{p} + \boldsymbol{\pi})\, ds. \tag{99}$$

4.2.4 Discretization in Space: Finite Elements

The discretization in space of the problem is carried out with the help of the finite element method. We consider a partition of $[0, L]$ in N_e linear elements with constant length h. The basic idea consists in approximating $\mathcal{Q}$ by means of a finite-dimensional subspace $\mathcal{Q}_h$. In following, calculations are performed on the basis of a generic finite element.

(i) We consider first the translational part of the motion. The current position of φ is approximated by

$$\varphi_h(\zeta) = \frac{1}{2}(1-\zeta)\varphi_1 + \frac{1}{2}(1+\zeta)\varphi_2, \qquad \zeta \in [-1, 1], \tag{100}$$

where φ_1, φ_2 are the position vectors of the initial and final nodes of a generic element in the mesh (Hughes 1987). It is worth noting that $\varphi_h(\zeta)$ belongs to $\mathbb{R}^3$ for all $\zeta \in [-1, 1]$ and that

$$\frac{\partial \varphi_h}{\partial s}(\zeta) = \frac{\varphi_2 - \varphi_1}{h}, \qquad \zeta \in [-1, 1],$$

since $ds/d\zeta = 2/h$.

(ii) Since the rotation group is a nonlinear manifold, (100) cannot be applied. Instead, we adopt the following procedure proposed in (Sansour and Wagner 2003),

- Use the *Spurrier's algorithm* (Spurrier 1978) to represent the nodal values of the rotation tensors, $\{\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2\}$, in terms of unit quaternions[8] $\{\mathbf{q}_1, \mathbf{q}_2\}$.
- The nodal values of the rotation vectors $\{\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2\}$ are extracted from $\{\mathbf{q}_1, \mathbf{q}_2\}$ with the help of the procedure given in (Simo and Vu-Quoc 1986).
- $\boldsymbol{\Psi}_h(s)$ can be obtained by applying (100). Then, $\mathbf{q}_h(s)$ is computed from $\boldsymbol{\Psi}_h(s)$ following standard procedures (see e.g. Crisfield 1998, Chap. XVI).
- Finally, $\boldsymbol{\Lambda}_h(s)$ is obtained from $\mathbf{q}_h(s)$ by applying the classical relation between rotation tensors and quaternions.

The construction of a semi-discrete counterpart of the Lagrangian function (95a) is as follows. The semi-discrete kinetic energy $\mathsf{K}^h : \mathsf{T}\mathcal{Q}_h \times \mathbb{R} \to \mathbb{R}$ and the internal and external components of the potential energy $\mathsf{U}^{h,\text{int}}, \mathsf{U}^{h,\text{ext}} : \mathcal{Q}_h \times \mathbb{R} \to \mathbb{R}$ are given by

$$\begin{aligned}
\mathsf{K}^h &= \frac{h}{4}\left(\mathcal{A}_\rho(\|\dot{\varphi}_1\|^2 + \|\dot{\varphi}_2\|^2\right) - \operatorname{Tr}\Big[\widetilde{\mathbf{V}}_1\mathbf{E}_\rho\widetilde{\mathbf{V}}_1 + \widetilde{\mathbf{V}}_2\mathbf{E}_\rho\widetilde{\mathbf{V}}_2\Big], && \text{(101a)}\\
\mathsf{U}^{h,\text{int}} &= h\,\psi\left(\boldsymbol{\Gamma}_h(0), \boldsymbol{\Omega}_h(0)\right), && \text{(101b)}\\
\mathsf{U}^{h,\text{ext}} &= h\,\mathrm{w}_{\text{ext}}\left(\varphi_h(0), \boldsymbol{\Lambda}_h(0)\right), && \text{(101c)}
\end{aligned}$$

where the explicit dependence on time has been omitted to simplify notation and

$$\begin{aligned}
\boldsymbol{\Gamma}_h(0) &= \boldsymbol{\Lambda}_h^t(0)\frac{\partial \varphi_h}{\partial s}(0) - \mathbf{E}_1,\\
\boldsymbol{\Omega}_h(0) &= \operatorname{axial}\left[\widetilde{\boldsymbol{\Omega}}_h(0)\right] = \operatorname{axial}\left[\boldsymbol{\Lambda}_h^t(0)\frac{\partial \boldsymbol{\Lambda}_h}{\partial s}(0)\right],
\end{aligned}$$

[8] Theoretical aspect about unit quaternions can be consulted, e.g., in (Mcrobie and Lasenby 1999; Crisfield 1998).

The semi-discrete Lagrangian is then constructed as

$$\mathsf{L}^h = \mathsf{K}^h - \mathsf{U}^{h,\mathrm{int}} - \mathsf{U}^{h,\mathrm{ext}}.$$

The application of Hamilton's principle allows to obtain the corresponding EL equations on the nodes of the mesh.

4.2.5 Discretization in Time: Variational Integrators

We denote by

$$(\boldsymbol{\varphi}_i^k, \boldsymbol{\Lambda}_i^k), \qquad i = 1, 2, \tag{102}$$

to the approximation of the nodal variables $\left(\boldsymbol{\varphi}_i(t^k), \boldsymbol{\Lambda}_a^k(t^k)\right)$.

The nodal value of the translational velocity is approximated by

$$\dot{\boldsymbol{\varphi}}_i(\tau) \approx \frac{\boldsymbol{\varphi}_i^{k+1} - \boldsymbol{\varphi}_i^k}{\Delta t}, \qquad i = 1, 2, \quad \tau \in (t^k, t^{k+1}), \tag{103}$$

and time derivative of the rotational tensor by

$$\dot{\boldsymbol{\Lambda}}_i(\tau) \approx \frac{\boldsymbol{\Lambda}_i^{k+1} - \boldsymbol{\Lambda}_h^k}{\Delta t}, \qquad i = 1, 2, \quad \tau \in (t^k, t^{k+1}). \tag{104}$$

The fully discrete (in space and time) counterpart of the kinetics energy is obtained by replacing $\dot{\boldsymbol{\varphi}}_i$ by (103) and $\dot{\boldsymbol{\Lambda}}_i$ by (104) in (101a) to obtain

$$\mathsf{K}_d^h\left(\boldsymbol{\Phi}_h^0, \boldsymbol{\Phi}_h^1\right) = \frac{h}{2(\Delta t)^2} \sum_{i=1}^{2} \left(\frac{\mathcal{A}_\rho}{2} \|\boldsymbol{\varphi}_i^{k+1} - \boldsymbol{\varphi}_i^k\|^2 + \mathrm{Tr}\left[\left(\mathbf{I} - \boldsymbol{\Lambda}_i^{(k)t} \boldsymbol{\Lambda}_i^{k+1}\right) \mathbf{E}_\rho\right]\right). \tag{105}$$

Correspondingly, the discrete potential energy, $\mathsf{U}_d : \mathcal{Q}_h \to \mathbb{R}$, is given by

$$\mathsf{U}_d^h(\boldsymbol{\Phi}_h^k) = h \,\left(\psi\left(\boldsymbol{\Gamma}_h^k(0), \widetilde{\boldsymbol{\Omega}}_h^k(0)\right) - \mathrm{w}_{\mathrm{ext}}\left(\boldsymbol{\varphi}_h^k(0), \boldsymbol{\Lambda}_h^k(0)\right)\right). \tag{106}$$

Consider $\alpha \in [0, 1]$ and apply the generalized trapezoidal rule (Marsden and West 2001) to construct a discrete Lagrangian,

$$\mathsf{L}_d^{h,\alpha}\left(\boldsymbol{\Phi}_h^0, \boldsymbol{\Phi}_h^1\right) = \Delta t \left(\mathsf{K}_d^h\left(\boldsymbol{\Phi}_h^0, \boldsymbol{\Phi}_h^1\right) - \alpha \mathsf{U}_d^h\left(\boldsymbol{\Phi}_h^0\right) - (1-\alpha)\mathsf{U}_d^h\left(\boldsymbol{\Phi}_h^1\right)\right). \tag{107}$$

The application of the discrete Hamilton's principle yields to the following DEL equations

$$\mathrm{m}_i \frac{\varphi_i^{k+1} - 2\varphi_i^k - \varphi_i^{k-1}}{(\Delta t)^2} = \mathbf{s}_i^k + \mathbf{S}_i^k,$$

$$\frac{\mathbf{R}_i^{(k)\mathrm{t}} - \mathbf{R}_i^{(k-1)\mathrm{t}} - \mathbf{R}_i^k + \mathbf{R}_i^{(k-1)}}{(\Delta t)^2} = \widetilde{\mathbf{h}}_i^k + \widetilde{\mathbf{H}}_i^k, \qquad i = 1, 2,$$

where

$$\mathrm{m}_i = \frac{h}{2}\mathcal{A}_\rho, \tag{109a}$$

$$\mathbf{R}_i^k = \frac{h}{2}\mathbf{J}_{\rho(i)}^k \boldsymbol{\Lambda}_{\mathrm{in}(i)}^{(k)\mathrm{t}}, \tag{109b}$$

$$\mathbf{s}_i^k = (-1)^i \mathbf{n}_h^k(0), \tag{109c}$$

$$\widetilde{\mathbf{h}}_i^k = (-1)^{i+1}\widetilde{\mathbf{m}}_h^k(0) + \frac{(\widetilde{\varphi}_2^k - \widetilde{\varphi}_1^k)}{2}\widetilde{\mathbf{n}}_h^k(0) - \widetilde{\mathbf{n}}_h^k(0)\frac{(\widetilde{\varphi}_2^k - \widetilde{\varphi}_1^k)}{2}, \tag{109d}$$

$$\mathbf{S}_i^k = -\frac{h}{2}\mathbf{N}_h^k(0) \quad \text{and} \quad \widetilde{\mathbf{H}}_i^k = \frac{h}{2}\widetilde{\mathbf{M}}_h^k(0), \tag{109e}$$

with $\mathbf{J}_{\rho(i)}^k = \boldsymbol{\Lambda}_i^k \mathbf{E}_\rho \boldsymbol{\Lambda}_i^{(k)\mathrm{t}}$ and $\boldsymbol{\Lambda}_{\mathrm{in}(i)}^k = \boldsymbol{\Lambda}_i^{(k)\mathrm{t}} \boldsymbol{\Lambda}_i^{k+1}$.

For the translational part of the motion, the discrete Legendre transform allows to obtain

$$\mathbf{p}_i^k = \mathrm{m}_i \frac{\varphi_i^{k+1} - \varphi_i^k}{\Delta t} + \alpha \Delta t \left(\mathbf{s}_i^k + \mathbf{S}_i^k\right), \tag{110a}$$

$$\mathbf{p}_i^{k+1} = \mathrm{m}_i \frac{\varphi_i^{k+1} - \varphi_i^k}{\Delta t} - (1-\alpha)\Delta t \left(\mathbf{s}_i^{k+1} + \mathbf{S}_i^{k+1}\right), \quad i = 1, 2. \tag{110b}$$

For the rotational part the following relation holds,

$$\widetilde{\boldsymbol{\Pi}}_i^k = \frac{\mathbf{R}_i^{(k)\mathrm{t}} - \mathbf{R}_i^k}{\Delta t} - \alpha \Delta t \left(\widetilde{\mathbf{h}}_i^k + \widetilde{\mathbf{H}}_i^k\right), \tag{111a}$$

$$\widetilde{\boldsymbol{\Pi}}_i^{k+1} = \frac{\mathbf{R}_i^{(k)\mathrm{t}} - \mathbf{R}_i^k}{\Delta t} + (1-\alpha)\Delta t \left(\widetilde{\mathbf{h}}_i^{k+1} + \widetilde{\mathbf{H}}_i^{k+1}\right), \quad i = 1, 2. \tag{111b}$$

Subtracting (110a) from (110b) and (111a) from (111b), yields to the following relations:

$$\mathbf{p}_i^{k+1} = \mathbf{p}_i^k - \Delta t \left(\mathbf{s}_i^{k+\alpha} + \mathbf{S}_i^{k+\alpha}\right), \tag{112a}$$

$$\widetilde{\boldsymbol{\Pi}}_i^{k+1} = \widetilde{\boldsymbol{\Pi}}_i^k + \Delta t \left(\widetilde{\mathbf{h}}_i^{k+\alpha} + \widetilde{\mathbf{H}}_i^{k+\alpha}\right), \qquad i = 1, 2, \tag{112b}$$

where $(\bullet)_a^{k+\alpha} = (1-\alpha)(\bullet)_a^{k+1} + \alpha(\bullet)_a^k$. A detailed deduction of (110a), (110b), (111a), and (111b) may be found in (Mata 2015).

4.2.6 Solution Procedure

The updating procedure for the nodal values of the configuration variables is as follows.

- Since mass matrix is diagonal, it is possible to update the nodal positions explicitly according to
$$\boldsymbol{\varphi}_i^{k+1} = \boldsymbol{\varphi}_i^k + \frac{\Delta t}{\mathrm{m}_i}\left(\mathbf{p}_i^k - \alpha\Delta t\left(\mathbf{s}_i^k + \mathbf{S}_i^k\right)\right). \tag{113}$$

- Equation (111a) yields to
$$\boldsymbol{\Lambda}_{\text{in}(i)}^k \mathbf{J}_{\rho(i)}^k - \mathbf{J}_{\rho(i)}^k \boldsymbol{\Lambda}_{\text{in}(i)}^{k(\text{t})} = \frac{2\Delta t}{h}\left(\widetilde{\boldsymbol{\Pi}}_i^k + \alpha\Delta t\left(\widetilde{\mathbf{h}}_i^k + \widetilde{\mathbf{H}}_i^k\right)\right). \tag{114}$$

  To determine $\boldsymbol{\Lambda}_{\text{in}(i)}^k$ we follow the method proposed in (Lee et al. 2007). We parametrize the incremental rotation tensor in terms of an incremental rotation vector as
$$\boldsymbol{\Lambda}_{\text{in}(i)}^k = \exp\left[\text{skew}\left(\boldsymbol{\theta}_i^k\right)\right].$$

  Then, replacing the above expression in (114) the following nonlinear system of equations is obtained
$$\frac{\sin\theta_i^k}{\theta_i^k}\left(\mathbf{i}_{\rho(i)}^k \boldsymbol{\theta}_i^k\right) + \frac{1-\cos\theta_i^k}{(\theta_i^k)^2}\left(\boldsymbol{\theta}_i^k \times \mathbf{i}_{\rho(i)}^k \boldsymbol{\theta}_i^k\right) = \mathbf{Y}_i^k, \tag{115}$$

  where $\theta = \|\boldsymbol{\theta}\|$ and $\mathbf{Y}_i^k$ is the axial vector of the right-hand side of (114).The system (115) is solved with the help of the Newton–Raphson scheme.

  Having obtained $\boldsymbol{\theta}_i^k$, we update the nodal rotation tensors according to
$$\boldsymbol{\Lambda}_a^{k+1} = \exp[\widetilde{\boldsymbol{\theta}}_i^k]\boldsymbol{\Lambda}_i^k. \tag{116}$$

- Finally, the nodal values of the momenta are updated with the help of (112a) and (112b).

Remark We formulated an explicit method to update translational part of the configuration variables, (see (113)). However, to update the rotational part a nonlinear system of equations has to be solved iteratively in every node of the mesh. ■

4.2.7 Properties of the Resulting Scheme

The resulting integration scheme enjoys several properties which are described in following.

- The discretization in space allows to formulate an explicit time integrator since mass matrix is diagonal and positive. Moreover, the order of accuracy with the mesh size does not result affected as explained in (Mata 2015; Cohen et al. 2001).
- Fixing $h > 0$, the order of accuracy of an algorithm in position-momentum form coincides with the order of accuracy with which the discrete Lagrangian approximates the exact discrete Lagrangian (if some standard smooth conditions hold). See (Marsden and West 2001). In our case, second order of accuracy is obtained if $\alpha = \frac{1}{2}$ since the resulting method is symmetric.
- Variational methods automatically conserve a discrete analogue of the symplectic two-form (see, e.g., Lew et al. 2003; Marsden and West 2001 for a proof). Moreover, the discrete total energy

$$\mathsf{H}_d^k = \frac{1}{2}\sum_{a=1}^{n_t}\left(\mathrm{m}_a^{-1}\mathbf{p}_a^k \cdot \mathbf{p}_a^k + \mathbf{I}_{\rho(a)}^{-1}\mathbf{\Pi}_a^k \cdot \mathbf{\Pi}_a^k\right) + \mathsf{U}_d\left(\boldsymbol{\varphi}_1^k, ..., \boldsymbol{\varphi}_{n_d}^k, \mathbf{\Lambda}_1^k, ..., \mathbf{\Lambda}_{n_d}^k\right), \tag{117}$$

  remains $\mathcal{O}\left((\Delta t)^2\right)$ close to the exact value for exponentially long periods of time if a small enough $\Delta t > 0$ is provided (Hairer et al. 2006). In the above equation n_t is the total number of nodes in the mesh.
- $\mathsf{L}_d^{h,\alpha}$ results to be invariant under the action of rigid body translations and rotations. Therefore, according to a discrete version of Noether's theorem the discrete versions of the total linear momentum and total angular velocity

$$\mathbf{L}_{t,d} = \sum_{a=1}^{n_t}\mathbf{p}_a^k \quad \text{and} \quad \mathbf{J}_{t,d} = \sum_{a=1}^{n_t}(\widetilde{\mathbf{\Pi}}_a^k + \boldsymbol{\varphi}_a^k \times \mathbf{p}_a^k), \tag{118}$$

  are invariants of the discrete trajectories.
- Regarding to the stability limit, we follow (Lew et al. 2003) choosing Δt as a fraction of the critical time step length imposed by the Courant condition.

4.2.8 Numerical Example: Elastic Ring

In this example, we use the rod model to simulate the dynamics of the elastic ring shown in Fig. 19. The applied load are

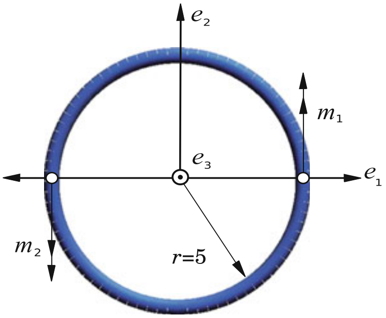
$$\boldsymbol{m}_1 t) = m(t)\mathbf{e}_2 \quad \text{and} \quad \boldsymbol{m}_2(t) = -\boldsymbol{m}_1(t),$$

where

$$m(t) = \begin{cases} (160/3)t, & 0 \le t \le 1.5 \\ 0, & t > 0 \end{cases}. \tag{119}$$

Since the applied moments are self-equilibrated, both the total linear momentum and the total angular momentum are equal to zero and exactly conserved during the free-fly phase of the motion. The mechanical properties of the cross-section are:

Fig. 19 Elastic ring



$E\mathcal{A} = G\mathcal{A} = 3 \times 10^4$, $E\mathrm{I} = 7 \times 10^3$, $\mathcal{A}_{\rho_0} = 1$ and $\mathrm{I}_{\rho_0} = 10$. The ring is discretized using 80 linear elements. The simulation is carried out during 4.5 s with a time step length $\Delta t = 0.0015$ s. The system exhibits drastic changes in the configuration during the motion. Figure 20 shows a sequence of snapshots of the motion in the $\mathbf{e}_1$-$\mathbf{e}_2$ plane. Finally, the time evolution of total energy is shown in Fig. 21 from which it is possible to verify the excellent long-term energy behavior of the algorithm.

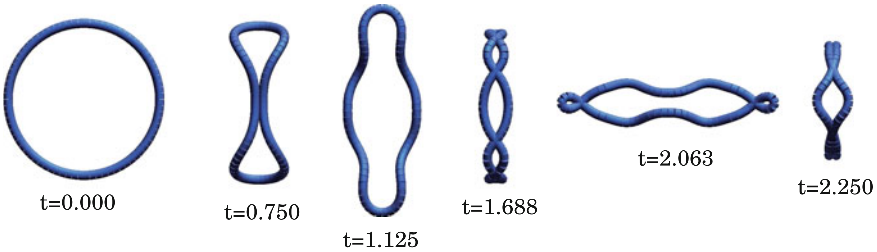


Fig. 20 Sequence of snapshots of the motion in the $\mathbf{e}_1$-$\mathbf{e}_2$ plane. Note that self-contact is not prevented in the present form of the algorithm

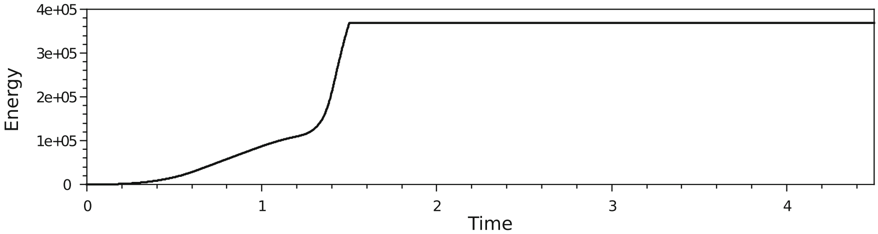


Fig. 21 Time evolution of the total energy

A (Smooth) Manifolds and Lie Groups

In this section, we briefly introduce the concepts of smooth n-dimensional manifolds and Lie groups in order to use them to make a proper introduction to finite rotations in the following section. The interested reader may consult standard textbooks such as, e.g., (Abraham et al. 1988, Arnold 1989, Mishchenko and Fomenko 1988).

A.1 Smooth n-Manifolds

A *smooth n–manifold* or manifold modeled in $\mathbb{R}^n$ is a set $\mathcal{M}$ such that:

- For each element $P \in \mathcal{M}$ there exists a subset $\mathcal{U}$ of $\mathcal{M}$ containing P and an one-to-one mapping called a *chart* or *coordinate system*, $\{x^\alpha\}_{\alpha=1,2,\ldots,n}$, from $\mathcal{U}$ onto an open set $\mathcal{V} \in \mathbb{R}^n$; x^α denote the components of this mapping ($\alpha = 1, 2, \ldots, n$).
- If x^α and $\overline{x}^\alpha$ are two of such mappings, the change of coordinate functions $\overline{x}^\alpha(x^1, \ldots, x^n)$ are C^∞ (*i.e.* it is continuously differentiable as many times as required).

Tangent Space

Let $\mathcal{M} \subset \mathbb{R}^n$ be an open set (manifold) and let $P \in \mathcal{M}$. The *tangent space* to $\mathcal{M}$ at P is simply the vector space $\mathbb{R}^n$ regarded as vectors emanating from P; this tangent space is denoted $T_P\mathcal{M}$.

A.2 Lie Groups

A *Lie group* is a smooth n–dimensional manifold $\mathcal{M}^n$ endowed with the following two smooth mappings:

(i) Multiplication:

$$\begin{aligned} \mathscr{F}_\alpha : \mathcal{M}^n \times \mathcal{M}^n &\to \mathcal{M}^n \\ (\mathbf{u}, \mathbf{v}) &\mapsto \mathscr{F}_\alpha(\mathbf{u}, \mathbf{v}) = \mathbf{u} \odot \mathbf{v}. \end{aligned}$$

where $\odot$ is used to indicate an abstract operation (multiplication) between elements of the manifold $\mathcal{M}^n$.

(ii) Construction of the inverse element:

$$\begin{aligned} \mathscr{F}_\beta : \mathcal{M}^n &\to \mathcal{M}^n \\ \mathbf{u} &\mapsto \mathscr{F}_\beta(\mathbf{u}) = \mathbf{u}^{-1}. \end{aligned}$$

Moreover, a Lie group posses a marked point $\mathbf{e} \in \mathcal{M}^n$ (the identity) which satisfies together with $\mathscr{F}_\alpha$ and $\mathscr{F}_\beta$ the following relations:

- $\boldsymbol{x}_1 \odot (\boldsymbol{x}_2 \odot \boldsymbol{x}_3) = (\boldsymbol{x}_1 \odot \boldsymbol{x}_2) \odot \boldsymbol{x}_3, \quad \text{for all} \quad \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in \mathcal{M}^n.$
- $\mathbf{e} \odot \boldsymbol{x}_1 = \boldsymbol{x}_1 \odot \mathbf{e} = \boldsymbol{x}, \qquad \boldsymbol{x}, \mathbf{e} \in \mathcal{M}^n.$
- $\boldsymbol{x} \odot \boldsymbol{x}^{-1} = \boldsymbol{x}^{-1} \odot \boldsymbol{x} = \mathbf{e}, \qquad \boldsymbol{x}, \boldsymbol{x}^{-1}, \mathbf{e} \in \mathcal{M}^n.$

Lie Algebra

The *Lie algebra*, $\mathfrak{g}$, of the Lie group G is given by its tangent vector space at the identity, $\mathfrak{g} = T_{\mathbf{e}}G$, equipped with a bilinear, skew–symmetric brackets operator $[\cdot, \cdot]$ satisfying the following relations (Dubrokin et al. 2000; Mishchenko and Fomenko 1988):

(i) Jacobi's identity:

$$[\boldsymbol{x}_a, [\boldsymbol{x}_b, \boldsymbol{x}_c]] + [\boldsymbol{x}_b, [\boldsymbol{x}_c, \boldsymbol{x}_a]] + [\boldsymbol{x}_c, [\boldsymbol{x}_a, \boldsymbol{x}_b]] = 0 \quad \text{for all} \quad \boldsymbol{x}_a, \boldsymbol{x}_b, \boldsymbol{x}_c \in \mathfrak{g}.$$

(ii) Skew-symmetry:

$$[\boldsymbol{x}_a, \boldsymbol{x}_b] = -[\boldsymbol{x}_b, \boldsymbol{x}_a] \quad \text{for all} \quad \boldsymbol{x}_a, \boldsymbol{x}_b \in \mathfrak{g},$$

where the *Lie brackets* are given by $[\boldsymbol{x}_a, \boldsymbol{x}_b] = \boldsymbol{x}_a \odot \boldsymbol{x}_b - \boldsymbol{x}_b \odot \boldsymbol{x}_a$.

B Finite Rotations

This section provides a brief introduction to finite rotations and to the rotational motion. We restrict the survey to such concepts that are used through the sections of the chapter. A more extensive review can be found, e.g., in (Argyris 1982; Argyris and Poterasu 1993; Atluri and Cazzani 1995; Bauchau and Trainelli 2003).

B.1 Noncommutative Rotations

Consider the *rotation vector*

$$\hat{\boldsymbol{\phi}} = (\phi_{xx}, \phi_{yy}, \phi_{zz}) = (\pi/2, \pi/2, \pi/2).$$

Figure 22 shows that the order in which we apply the components of the rotation vector on a rigid body (in this case a rigid box) affects its final configuration in space. Therefore, $\hat{\boldsymbol{\phi}}$ can not be used to represent uniquely a rotation in space. Or, in

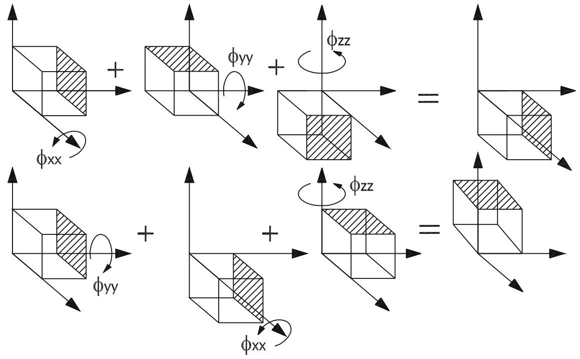


Fig. 22 Noncommutativity of the components of the *rotation vector* $\hat{\phi} = [\phi_{xx}, \phi_{yy}, \phi_{zz}]$

other words, rotations are not elements of a vector space (Jeleniĉ and Crisfield 1999; Simo and Vu-Quoc 1986).

Alternatively, we may think of a rotation $\boldsymbol{\beta}$ as a linear application from the Euclidean vector space, $\mathbb{E}^3$, to itself. Therefore, when $\boldsymbol{\beta}$ is applied to a vector $\boldsymbol{u} \in \mathbb{E}^3$, the result is a new vector $\boldsymbol{v} = \boldsymbol{\beta u}$ conserving the original length. Consider the set

$$\mathfrak{R} = \{\boldsymbol{\beta} : \mathbb{E}^3 \to \mathbb{E}^3 \mid \boldsymbol{\beta} \text{ is a rotation}\},$$

and define the *sum of rotations* as

$$(\boldsymbol{\beta}_a \otimes \boldsymbol{\beta}_b)(\boldsymbol{x}) = \boldsymbol{\beta}_a\left(\boldsymbol{\beta}_b(\boldsymbol{x})\right) \qquad \boldsymbol{\beta}_a, \boldsymbol{\beta}_b \in \mathfrak{R}, \quad \boldsymbol{x} \in \mathbb{E}^3. \tag{120}$$

Clearly $\boldsymbol{\beta}_a \otimes \boldsymbol{\beta}_b \in \mathfrak{R}$ is a *compound rotation* applied on $\boldsymbol{x}$.

The set $\mathfrak{R}$ equipped with the operation $\otimes$ posses the algebraic structure of *noncommutative group* (Bauchau and Trainelli 2003; Mäkinen and Marjamäki 2005; Mäkinen 2004; Mishchenko and Fomenko 1988) and enjoys the following properties:

(i) Associativity:

$$\boldsymbol{\beta}_a \otimes (\boldsymbol{\beta}_b \otimes \boldsymbol{\beta}_c) = (\boldsymbol{\beta}_a \otimes \boldsymbol{\beta}_b) \otimes \boldsymbol{\beta}_c, \quad \text{for all} \quad \boldsymbol{\beta}_a, \boldsymbol{\beta}_b, \boldsymbol{\beta}_c \in \mathfrak{R}.$$

(ii) There exists a unique identity element $\boldsymbol{i} \in \mathfrak{R}$ such that

$$\boldsymbol{\beta} \otimes \boldsymbol{i} = \boldsymbol{i} \otimes \boldsymbol{\beta} = \boldsymbol{\beta}, \quad \text{for all} \quad \boldsymbol{\beta} \in \mathfrak{R}.$$

(iii) For each $\boldsymbol{\beta} \in \mathfrak{R}$ there exists a unique element belonging to $\mathfrak{R}$ called the inverse of $\boldsymbol{\beta}$ and denoted by $\boldsymbol{\beta}^{-1}$ such that

$$\boldsymbol{\beta}^{-1} \otimes \boldsymbol{\beta} = \boldsymbol{\beta} \otimes \boldsymbol{\beta}^{-1} = \boldsymbol{i}.$$

(iv) The operation $\otimes$ is, in general, noncommutative, *i.e.* ,

$$\boldsymbol{\beta}_a \otimes \boldsymbol{\beta}_b(\boldsymbol{x}) \neq \boldsymbol{\beta}_b \otimes \boldsymbol{\beta}_a(\boldsymbol{x}).$$

for all $\boldsymbol{\beta}_a, \boldsymbol{\beta}_b \in \mathfrak{R}$ and $\boldsymbol{x} \in \mathbb{E}^3$.

The following result is fundamental:

Theorem *The group $\mathfrak{R}$ is isomorphic to the set composed by all the real and orthogonal matrices of order 3, with determinant equal to 1.*

Proof See, e.g., (Pérez-Morán 2005) and references therein ■

This theorem allows to identify each finite rotation with an orthogonal *rotation tensor* belonging to the *special orthogonal group*, $SO(3)$, defined as

$$SO(3) = \left\{ \boldsymbol{\Lambda} \in \mathbf{M}^{3\times3} \quad | \quad \boldsymbol{\Lambda}^t \boldsymbol{\Lambda} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^t = \mathbf{I}; \quad \det[\boldsymbol{\Lambda}] = 1 \right\}, \tag{121}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{M}^{3\times3}$ is the set composed by all the 3 × 3 matrices with real coefficients. It is not difficult to see that $SO(3)$ also has the structure of a smooth differentiable manifold (Dubrokin et al. 2000). See Sect. A.1. Moreover, under the usual matrix multiplication, it has the structure of a Lie group. See Sect. A.2.

The components of a rotation tensor depend on the reference frame adopted and thus to compose two rotations $\boldsymbol{\Lambda}_a, \boldsymbol{\Lambda}_b \in SO(3)$ two situations can happen:

(i) *Spatial description of rotations*. In this case the components of $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$ are expressed in terms of a fixed reference frame (Argyris 1982). The rotation tensor representing the result of applying $\boldsymbol{\Lambda}_b$ after $\boldsymbol{\Lambda}_a$ is obtained as

$$\boldsymbol{\Lambda}_b \circ \boldsymbol{\Lambda}_a = \boldsymbol{\Lambda}_b \boldsymbol{\Lambda}_a \in SO(3),$$

Therefore, the *inverse multiplicative rule* for rotation tensors applies.

(ii) *Material description of rotations*. In the second case, $\boldsymbol{\Lambda}_a$ moves the reference frame and therefore, the components of $\boldsymbol{\Lambda}_b$ are expressed in an updated reference frame. Then, we have that

$$\boldsymbol{\Lambda}_b \circ \boldsymbol{\Lambda}_a = \boldsymbol{\Lambda}_a \boldsymbol{\Lambda}_b \in SO(3).$$

B.2 Parametrization of $SO(3)$

The rotational motion can be described by a means of a trajectory on $SO(3)$. Therefore, it can not be described trivially by using standard coordinates such as those employed for trajectories belonging to a linear space. Rotations may be parametrized using suitable charts which are inherently not global and/or singular. Over the years, numerous techniques have been developed to cope with the description of rotational motion (Pérez-Morán 2005; Simo and Vu-Quoc 1986; Bauchau and Choi 2003; Milenkovic and Milenkovic 1997; Stuelpnagel 1964; Grassia 1998; Cottingham and Doyle 2001; Innocenti and Paganelli 2006). All these techniques show certain balance between advantages and drawbacks when compared each to other. In the following, we describe a (minimal) parametrization of rotation tensors in terms of rotation vectors. It is based on the following result:

Fundamental Theorem of Euler

The general displacement of a rigid body or vector, with one point fixed is a rotation about some axis which passes through that point.

Proof See (Crisfield 1998, Vol 2.) ■

Basically, the above theorem shows that the rotational motion is completely described by a unit vector $\hat{\mathbf{e}} \in \mathbb{R}^3$ defining an axis of rotation in space and a *rotation angle* of magnitude $\theta \in [0, 2\pi]$. Moreover, the corresponding rotation tensor is expressed according to the *Rodrigues's formula*:

$$\mathbf{\Lambda} = \mathbf{I} + \frac{\sin\theta}{\theta}\widetilde{\boldsymbol{\theta}} + \frac{(1-\cos\theta)}{\theta^2}\widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}} = \mathbf{I} + \sin\theta\widetilde{\mathbf{e}} + (1-\cos\theta)\widetilde{\mathbf{e}}\widetilde{\mathbf{e}}. \tag{122}$$

where we use $\widetilde{\boldsymbol{u}}$ to denote the skew-symmetric tensor obtained from the vector $\boldsymbol{u} \in \mathbb{R}^3$.

B.3 Tangent Spaces

Consider $\mathbf{\Lambda} \in SO(3)$. The variations of $\mathbf{\Lambda}\mathbf{\Lambda}^t$ and $\mathbf{\Lambda}^t\mathbf{\Lambda}$ are given by

$$\delta(\mathbf{\Lambda}\mathbf{\Lambda}^t) = \delta\mathbf{\Lambda}\mathbf{\Lambda}^t + \mathbf{\Lambda}\delta\mathbf{\Lambda}^t = \widetilde{\boldsymbol{\phi}} + \widetilde{\boldsymbol{\phi}}^t = 0$$
$$\delta(\mathbf{\Lambda}^t\mathbf{\Lambda}) = \delta\mathbf{\Lambda}^t\mathbf{\Lambda} + \mathbf{\Lambda}^t\delta\mathbf{\Lambda} = \widetilde{\mathbf{\Phi}}^t + \widetilde{\mathbf{\Phi}} = 0,$$

from which it is possible to deduce that $\widetilde{\boldsymbol{\phi}}$ and $\widetilde{\mathbf{\Phi}}$ are skew-symmetric tensors. Moreover, we have that

$$\delta\mathbf{\Lambda} = \widetilde{\boldsymbol{\phi}}\mathbf{\Lambda} = \mathbf{\Lambda}\widetilde{\mathbf{\Phi}}. \tag{123}$$

Clearly $\delta\mathbf{\Lambda}$ belongs to the tangent space to $SO(3)$ at $\mathbf{\Lambda}$, $T_{\mathbf{\Lambda}}SO(3)$.

The tangent space at the identity forms the Lie algebra of $SO(3)$ and is denoted by

$$so(3) = T_{\mathbf{I}}SO(3).$$

From (123) it is possible to see that $so(3)$ corresponds to the linear space of skew-symmetric tensors of the form

$$\widetilde{\boldsymbol{\theta}} = \begin{bmatrix} 0 & -\theta_3 & \theta_2 \\ \theta_3 & 0 & -\theta_1 \\ -\theta_2 & \theta_1 & 0 \end{bmatrix} = \text{skew}[\boldsymbol{\theta}]. \tag{124}$$

Since $so(3)$ is isomorphic to $\mathbb{R}^3$, every $\widetilde{\boldsymbol{\theta}} \in so(3)$ can be represented by a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3$. Moreover, $\widetilde{\boldsymbol{\Phi}} = \boldsymbol{\Lambda}^t \widetilde{\phi} \boldsymbol{\Lambda}$ and $\boldsymbol{\Phi} = \boldsymbol{\Lambda}^t \phi$.

The exponential map

$$\begin{aligned} \exp[\bullet] : so(3) &\to SO(3) \\ \widetilde{\boldsymbol{\theta}} &\mapsto \exp\left[\widetilde{\boldsymbol{\theta}}\right] = \boldsymbol{\Lambda}\left(\widetilde{\boldsymbol{\theta}}\right), \end{aligned} \tag{125}$$

allows to parametrize any rotation tensor in terms of an element of $so(3)$. See (122).

C Quaternions

The (minimal) vectorial parametrization of $SO(3)$ shows some limitations due to the fact that the exponential map is not a bijective application for angles greater than π (Pérez-Morán 2005; Simo and Vu-Quoc 1988; Crisfield 1998). The problem can be avoided if *unit quaternions* are used. A unit quaternion is defined using four parameters, q_0– q_3, so that:

$$\mathbf{q}_\theta = \cos(\theta/2) + \sin(\theta/2)\,\hat{\mathbf{e}} = \begin{bmatrix} \mathbf{q} \\ q_0 \end{bmatrix} = \begin{bmatrix} \sin\left(\frac{\theta}{2}\right)\hat{\mathbf{e}} \\ \cos\left(\frac{\theta}{2}\right) \end{bmatrix}, \tag{126}$$

where θ and $\hat{\mathbf{e}}$ are an angle of rotation and an axis of rotation, respectively. It is possible to see that $|\mathbf{q}_\theta| = 1$.

A rotation tensor $\boldsymbol{\Lambda} \in SO(3)$ is *uniquely* parametrized in terms of a unit quaternion $\mathbf{q}_\theta$ according to the formula:

$$\boldsymbol{\Lambda}(\mathbf{q}_\theta) = (\mathrm{q}_0^2 - \mathbf{q}\cdot\mathbf{q})\mathbf{I} + 2\mathbf{q}\otimes\mathbf{q} + 2\mathrm{q}_0\widetilde{\mathbf{q}}, \tag{127}$$

The quaternion compound rotation of the quaternions $\mathbf{q}_a = (a_0, \mathbf{a})$ and $\mathbf{q}_b = (b_0, \mathbf{b})$ is given by

$$\mathbf{q}_{ab} = \mathbf{q}_b\mathbf{q}_a = a_0 b_0 - \mathbf{a}\cdot\mathbf{b} + a_0\mathbf{b} + b_0\mathbf{a} - \mathbf{a}\times\mathbf{b}.$$

C.1 Normalized Quaternion from the Rotation Tensor

A general procedure for obtaining the rotation vector from the rotation tensor involves the computation of the Euler parameters, q_0–q_3. This can be achieved via the *Spurrier*'s algorithm (Crisfield 1998), which involves computing

$$a = \max\Big[\mathrm{Tr}[\boldsymbol{\Lambda}], \Lambda_{11}, \Lambda_{22}, \Lambda_{33}\Big]$$

where Tr[•] is the trace operator and

$$\begin{array}{lll}
\text{if} & a = \mathrm{Tr}(\boldsymbol{\Lambda}) \rightarrow & \begin{cases} \mathrm{q}_0 = \frac{1}{2}(1+a)^{\frac{1}{2}} \\ \mathrm{q}_i = (\Lambda_{kj} - \Lambda_{jk})/4q_0; \; i = 1, 3 \end{cases} \\
\text{else if} & a = \Lambda_{ii} \quad\; \rightarrow & \begin{cases} \mathrm{q}_i = (\frac{1}{2}a + \frac{1}{4}\big[1 - \mathrm{Tr}[\boldsymbol{\Lambda}]\big]) \\ \mathrm{q}_0 = \frac{1}{4}(\Lambda_{kj} - \Lambda_{jk})/q_i \\ \mathrm{q}_l = \frac{1}{4}(\Lambda_{li} + \Lambda_{il})/q_i; \qquad l = j, k \end{cases}
\end{array}$$

where i, j, k are a cyclic combinations of 1, 2, and 3.

References

Abraham, R., Marsden, J., & Ratiu, T. (1988). *Manifolds, Tensor Analysis, and Applications* (2nd ed., Vol. 75). Applied Mathematical Sciences. New York: Springer.

Ainsley, S., Vouga, E., Grinspun, E., & Tamstorf, R. (2012). Speculative parallel asynchronous contact mechanics. *ACM Transactions of Graphics*, *31*(6), 151:1–151:8.

Argyris, J. (1982). An excursion into large rotations. *Computer Methods in Applied Mechanics and Engineering*, *32*, 85–155.

Argyris, J., & Poterasu, V. F. (1993). Large rotations revisited application of lie algebra. *Computer Methods in Applied Mechanics and Engineering*, *103*, 11–42.

Armero, F., & Petoz, E. (1999). A new dissipative time-stepping algorithm for frictional contact problems: formulation and analysis. *Computer Methods in Applied Mechanics Engineering*, *179*, 151–178.

Armero, F., & Romero, I. (2001). On the formulation of high-frequency dissipative time-stepping algorithms for nonlinear dynamics. Part I: Low-order methods for two model problems and nonlinear elastodynamics. *Computer Methods in Applied Mechanics Engineering*, *190*, 2603–2649.

Armero, F., & Simo, J. C. (1992). A new unconditionally stable fractional step method for non-linear coupled thermomechanical problems. *International Journal for Numerical Methods in Engineering*, *35*(4), 737–766.

Arnold, V. I. (1989). *Mathematical methods of classical mechanics* (2nd ed., Vol. 60). Graduate texts in mathematics. New York: Springer.

Arnold, V. I. & Khesin, B. A. (1998). *Topological methods in hydrodynamics* (Vol. 125). Applied Mathematical Sciences. New York: Springer.

Arnold, V. I., Kozlov, V. V., & Neishtadt, A. I. (2006). *Mathematical aspects of classical and celestial mechanics* (3rd ed., Vol. 3)., Encyclopaedia of mathematical sciences. Berlin: Springer.

Atluri, S. N., & Cazzani, A. (1995). Rotations in computational solid mechanics. *Archives of Computational Methods in Engineering*, *2*(1), 49–138.

Bargmann, S., & Steinmann, P. (2008a). Modeling and simulation of first and second sound in solids. *International Journal of Solids and Structures*, *45*, 6067–6073.

Bargmann, S., & Steinmann, P. (2008b). An incremental variational formulation of dissipative and non-dissipative coupled thermoelasticity for solids. *Heat Mass Transfer*, *45*, 107–116.

Bathe, K. J. (1996). *Finite Element Procedures*. Prentice-Hall.

Bauchau, O., & Bottasso, C. (1999). On the design of energy preserving and decaying schemes for flexible, nonlinear multi-body systems. *Computer Methods in Applied Mechanics and Engineering*, *169*, 61–79.

Bauchau, O., & Trainelli, L. (2003). The vectorial parametrization of rotation. *Nonlinear Dynamics*, *32*, 71–92.

Bauchau, O. A., & Choi, J. I. (2003). The vector parameterization of motion. *Nonlinear Dynamics*, *33*, 165–188.

Bayliss, A., & Issacson, E. (1975). How to make your algorithm conservative. *American Mathematical Society*, *22*, A594–A595.

Betsch, P., & Uhlar, S. (2007). Energy-momentum conserving integration of multibody dynamics. *Multibody System Dynamics*, *17*, 243–289.

Betsch, P., Hesch, C., Sänger, N. & Uhlar, S. (2010). Variational integrators and energy-momentum schemes for flexible multibody dynamics. *Journal of Computational and Nonlinear Dynamics*, **5**(3):031001/1–11.

Bogolyubov, N. N. (1972). Part 2, chapter the model hamiltonian in superconductivity theory. *Particles and nuclei* (Vol. 1, pp. 1–52). US: Springer.

Borri, M., Bottasso, L., & Trainelli, L. (2001). Integration of elastic multibody system by invariant conserving/dissipating algorithms. I. Formulation. *Computer Methods in Applied Mechanics and Engineering*, *190*, 3669–3699.

Bou-Rabee, N., & Marsden, J. E. (2009). Hamilton-pontryagin integrators on lie groups part I: Introduction and structure-preserving properties. *Foundations of Computational Mathematics*, *9*(2), 197–219.

Bou-Rabee, N., & Owhadi, H. (2007). Stochastic variational partitioned runge-kutta integrators for constrained systems, arxiv.org/abs/0709.2222.

Bou-Rabee, N., & Owhadi, H. (2009). Stochastic variational integrators. *IMA Journal of Numerical Analysis*, *29*(2), 421–443.

Bou-Rabee, N., & Owhadi, H. (2010). Long-run accuracy of variational integrators in the stochastic context. *SIAM Journal on Numerical Analysis*, *48*(1), 278–297.

Cannarozzi, A. A., & Ubertini, F. (2001). A mixed variational method for linear coupled thermoelastic analysis. *International Journal of Solids and Structures*, *38*, 717–739.

Cano, B., & Sanz-Serna, J. M. (1988). Error growth in the numerical integration of periodic orbits by multistep methods, with application to reversible systems. *IMA Journal of Numerical Analysis*, *18*, 57–75.

Celledoni, E., & Owren, B. (2003). Lie group methods for rigid body dynamics and time integration on manifolds. *Computer Methods in Applied Mechanics and Engineering*, *192*(34), 421–438.

Celledoni, E., Marthinsen, H. & Owren, B. (2014). An introduction to lie group integrators - basics, new developments and applications. *Journal of Computational Physics*, **257**(Part B) 1040–1061.

Channell, P. J., & Scovel, C. (1990). Symplectic integration of hamiltonian systems. *Nonlinearity*, *3*, 231–259.

Chaturvedi, N. A., Lee, T., Leok, M., & McClamroch, N. H. (2011). Nonlinear Dynamics of the 3D Pendulum. *Journal of Nonlinear Science*, *21*(1), 3–32.

Chen, Y. (1990). A proof of the structure of the minimum-time control law of robotic manipulators using a hamiltonian formulation. *IEEE Robotics and Automation*, *6*(3), 388–393.

Chyba, M., Hairer, E., & Vilmart, G. (2009). The role of symplectic integrators in optimal control. *Optimal Control Applications and Methods*, *30*(4), 367–382.

Cirak, F., & West, M. (2005). Decomposition contact response (DCR) for explicit finite element dynamics. *International Journal for Numerical Methods in Engineering*, *64*(8), 1078–1110.

Clemente-Gallardo, J., & Scherpen, J. M. A. (2003). Relating lagrangian and hamiltonian formalisms of LC circuits. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, *50*(10), 1359–1363.

Cohen, G., Joly, P., Roberts, J. E., & Tordjman, N. (2001). Higher order triangular finite elements with mass lumping for the wave equation. *SIAM Journal on Numerical Analysis*, *38*(6), 2047–2078.

Coleman, B. D., & Noll, W. (1963). The thermodynamics of elastic materials with heat conduction and viscosity. *Archive for Rational Mechanics and Analysis*, *13*(3), 167–178.

Cottingham, W. N., & Doyle, D. D. (2001). The rotational dynamics of rigid bodies implemented with the cayley klein parametrization. *Molecular Physics*, *99*, 1839–1843.

Crisfield, M. A. (1998). *Non-linear finite element analysis of solids and structures* (Vol. 1, 2). Wiley.

De León, M., Marrero, J. C., & Martín De Diego, D. (2008). Some applications of semi-discrete variational integrators to classical field theories. *Qualitative Theory of Dynamical Systems*, *7*(1), 195–212.

Demoures, F., Gay-Balmaz, F., Kobilarov, M., & Ratiu, T. S. (2014). Multisymplectic lie group variational integrator for a geometrically exact beam in R^3. *Communications in Nonlinear Science and Numerical Simulation*, *19*(10), 3492–3512.

Demoures, F. M. A. (2012). *Lie group and lie algebra variational integrators for flexible beam and plate in R^3*. Ph.D thesis, École Polytechnique Fédérale de Lausanne.

Desbrun, M., Gawlik, E. S., Gay-Balmaz, F., & Zeitlin, V. (2014). Variational discretization for rotating stratified fluids. *Discrete and Continuous Dynamical Systems—Series A (DCDS-A)*, *34*(2), 477–509.

Dubrokin, B. A., Fomenko, A. T., & Nóvikov, S. P. (2000). *Geometría moderna. métodos y aplicaciones* (Vol. 1 & 2). Moscow: Mir, URSS.

Esposito, G., Marmo, G., & Sudarshan, G. (2004). *From classical to quantum mechanics: an introduction to the formalism foundations and applications*. New York: Cambridge University Press.

Faltinsen, S. (2000). Backward error analysis for Lie-group methods. *BIT Numerical Mathematics*, *40*(4), 652–670.

Feng, K., & Qin, M. (2010). *Symplectic Geometric Algorithms for Hamiltonian Systems*. Berlin: Zhejiang Publishing United Group Zhejiang Science and Technology Publishing House, Hangzhou and Springer.

Fetecau, R. C. (2003). *Variational methods for nonsmooth mechanics*. Ph.D thesis, California Institute of Technology, Pasadena, California, USA.

Fetecau, R. C., Marsden, J. E., & West, M. (2003a). Variational multisymplectic formulations of nonsmooth continuum mechanics. *Perspectives and problems in nonlinear science* (pp. 229–261). New York: Springer.

Fetecau, R. C., Marsden, J. E., Ortiz, M., & West, M. (2003b). Nonsmooth Lagrangian mechanics and variational collision integrators. *SIAM Journal on Applied Dynamical Systems*, *2*(3), 381–416.

Focardi, M., & Maria-Mariano, P. (2008). Convergence of asynchronous variational integrators in linear elastodynamics. *International Journal for Numerical Methods in Engineering*, *75*(7), 755–769.

Fong, W., Darve, E., & Lew, A. (2008). Stability of asynchronous variational integrators. *Journal of Computational Physics*, *227*, 8367–8394.

Gambar, K., & Markus, F. (1994). Hamilton-Lagrange formalism of nonequilibrium thermodynamics. *Physical Review E*, *50*(2), 1227–1231.

Gawlik, E. S., Mullen, P., Pavlov, D., Marsden, J. E., & Desbrun, M. (2011). Geometric, variational discretization of continuum theories. *Physica D: Nonlinear Phenomena*, *240*(21), 1724–1760.

Gay-Balmaz, F., Holm, D. D., & Ratiu, T. S. (2009). Variational principles for spin systems and the Kirchhoff rod. *The Journal of Geometric Mechanics (JGM)*, *1*(4), 417–444.

Ge, Z., & Marsden, J. E. (1988). Lie-poisson hamilton-jacobi theory and lie-poisson integrators. *Physics Letters A*, *133*(3), 134–139.

Gonzalez, O. (2000). Exact energy and momentum conserving algorithms for general models in nonlinear elasticity. *Computer Methods in Applied Mechanics and Engineering*, *190*, 1763–1783.

Grassia, F. S. (1998). Practical parameterization of rotations using the exponential map. *Journal of Graphics Tools*, *3*, 29–48.

Green, A., & Naghdi, P. (1991). A re-examination of the basic postulates of thermomechanics. *Proceedings: Mathematical and Physical Sciences*, *432*, 171–194.

Green, A., & Naghdi, P. (1995). A unified procedure for construction of theories of deformable media. I. classical continuum physics. *Mathematical and Physical Sciences*, *448*(1934), 335–356.

Green, A. E., & Naghdi, P. M. (1993). Thermoelasticity without energy dissipation. *Journal of Elasticity*, *31*(3), 189–208.

Grispun, E., Hirani, A., Desbrun, M. & Schröder, P. (2003). Discrete shells. In *Symposium on computer animation*, (pp. 62–67), San Diego, California.

Gross, M., & Betsch, P. (2006). An energy consistent hyprid space-time Galerkin method for nonlinear thermomechanical problems. *PAMM, Proceedings in Applied Mathematics and Mechanics*, *6*, 443–444.

Gross, M., Betsch, P. (2007). On deriving higher-order and energy-momentum-consistent time-stepping-schemes for thermo-viscoelastodynamics from a new hybrid space-time Galerkin method. In Bottasso, C. L., Masarati, P., & Trainelli, L., (Eds.), *Proceedings of the ECCOMAS Thematic Conference on Multibody Dynamics*, Milano, Italy: Politecnico di Milano.

Hairer, E., & Lubich, C. (1999). Invariant tori of dissipatively perturbed hamiltonian systems under symplectic discretization. *Applied Numerical Mathematics*, *29*(1), 57–71.

Hairer, E., & Wanner, G. (1996). *Solving ordinary differential equations II. Stiff and differential-algebraic problems* (Vol. 14). Springer Series in Computational Mathematics. Berlin: Springer.

Hairer, E., Norsett, S. P. & Wanner, G. (1993). *Solving ordinary differential equations I. Nonstiff problems*, (Vol 8). Springer Series in Computational Mathematics. Berlin: Springer.

Hairer, E., Lubich, C., & Wanner, G. (2003). Geometric numerical integration illustrated by the Störmer-Verlet method. *Acta Numerica*, *12*, 399–450.

Hairer, E., Lubich, C., & Wanner, G. (2006). *Geometric numerical integration. Structure preserving algorithms for ordinary differential equations*. Springer Series in Computational Mathematics. Springer.

Hall, J. & Leok, M. (2014a). Spectral variational integrators. *Numerische Mathematik*.

Hall, J. & Leok, M. (2014b). Spectral variational integrators. arXiv:1402.3327.

Harmon, D., Vouga, E., Smith, B., Tamstorf, R. & Grinspun, E. (2009). Asynchronous contact mechanics. In *SIGGRAPH'09 (ACM Transactions on Graphics)*, New York, USA: ACM, ISBN: 978-1-60558-726-4.

Holmes, M. H. (2007). *Introduction to numerical methods in differential equations* (Vol. 52). Texts in applied mathematics. New York: Springer.

Hughes, T. J. R. (1987). *The finite element method: linear static and dynamic finite element analysis*. Prentice Hall Inc.

Hutter, M., & Tervoort, T. A. (2007). Finite anisotropic elasticity and material frame indifference from a nonequilibrium thermodynamics perspective. *Journal of Non-Newtonian Fluid Mechanics*, *152*, 45–52.

Innocenti, C., & Paganelli, D. (2006). *Advances in robot kinematics mechanisms and motion*, chapter Determining the 3×3 rotation matrices that satisfy three linear equations in the direction cosines. Springer.

Iserles, A. (1997). *Foundations of computational mathematics*, chapter Numerical methods on (and off) manifolds, (pp. 180–189). Number 10208. Berlin: Springer.

Iserles, A., Munthe-Kaas, H. Z., Norsett, S. P., & Zanna, A. (2000). Lie-group methods. *Acta Numerica*, *9*, 215–365.

Jelenič, G. & Crisfield, M. A. (1999) Geometrically exact 3d beam theory: implementation of a strain-invariant finite element for static and dynamics. *Computer Methods in Applied Mechanics and Engineering*, *171*, 141–171.

Jiménez, F., Kobilarov, M., & Martín de Diego, M. (2013). Discrete variational optimal control. *Journal of Nonlinear Science*, *23*(3), 393–426.

Johnson, G., Leyendecker, S., & Ortiz, M. (2014). Discontinuous variational time integrators for complex multibody collisions. *International Journal for Numerical Methods in Engineering*, *100*(12), 871–913.

Kale, K. G., & Lew, A. J. (2006). Parallel asynchronous variational integrators. *International Journal for Numerical Methods in Engineering*, *70*(3), 291–321.

Kane, C., Marsden, J. E., & Ortiz, M. (1999). Symplectic-energy-momentum preserving variational integrators. *Journal of mathematical physics*, *40*(7), 3353–3371.

Kane, C., Marsden, J. E., Ortiz, M., & West, M. (2000). Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. *International Journal for Numerical Methods in Engineering*, *49*, 1295–1325.

Kapania, R. K., & Li, J. (2003). On a geometrically exact curved/twisted beam theory under rigid cross-section assumption. *Computational Mechanics*, *30*, 428–443.

Kasdin, N. J., Gurfil, P., & Kolemen, E. (2005). Canonical modelling of relative spacecraft motion via epicyclic orbital elements. *Celestial Mechanics and Dynamical Astronomy*, *92*.

Kern, D., Bär, S. & Groß, M. (2014). Variational integrators for thermomechanical coupled dynamic systems with heat conduction. *Proceedings in applied mathematics and mechanics, PAMM*, *14*(1), 47–48.

Kharevych, L., Weiwei, Y., Tong, Y., Kanso, E., Marsden, J. E., Schröder, P., & Desbrun, M. (2006). Geometric, variational integrators for computer animation. In *Eurographics/ACM SIGGRAPH Symposium on Computer Animation*.

Kirwan, A. D. (2008). Quantum and ecosystem entropies. *Entropy*, *10*, 58–70.

Kobilarov, M. (2014). *Multibody dynamics*, chapter Solvability of geometric integrators for multibody systems. (Vol. 35, pp. 145–174), Computational methods in applied sciences. Switzerland: Springer International Publishing.

Koon, W. S., Lo, M. W., Marsden, J. E. & Ross, S. D. (2011). *Dynamical systems, the three-body problem and space mission design*. Marsden Books.

Kraus, M. (2013). *Variational integrators in plasma physics*. Ph.D thesis, Technische Universität München.

Kuang, J., Leung, A. Y. T., & Tan, S. (2003). Hamiltonian and chaotic attitude dynamics of an orbiting gyrostat satellite under gravity-gradient torques. *Physica D: Nonlinear Phenomena*, *186*(1–2), 1–19.

Labudde, R. A., & Greenspan, D. (1976). Energy and momentum conserving methods of arbitrary order for the numerical integration of equations of motion part II. *Numerisch Mathematik*, *26*, 1–16.

Larsson, J. (1996). A new hamiltonian formulation for fluids and plasmas. part 3. multifluid electrodynamics. *Journal of Plasma Physics*, *55*(02), 279–300.

Lee, T., Leok, M., & McClamroch, N. H. (2007). Lie group variational integrators for the full body problem. *Computer Methods in Applied Mechanics and Engineering*, *196*(29–30), 2907–2924.

Lee, T., Leok, M., & McClamroch, N. H. (2009). Lagrangian mechanics and variational integrators on two-spheres. *International Journal for Numerical Methods in Engineering*, *79*(9), 1147–1174.

Leimkuhler, B., & Reich, S. (2005). *Simulating hamiltonian dynamics*. Cambridge Monographs on Applied and Computational Mathematics.

Leok, M. (2005). Generalized galerkin variational integrators. arXiv:math/0508360v1.

Leok, M., & Shingel, T. (2012a). General techniques for constructing variational integrators. *Frontiers of Mathematics in China*, *7*(2), 273–303.

Leok, M., & Shingel, T. (2012b). Prolongationcollocation variational integrators. *IMA Journal of Numerical Analysis*, *32*, 1194–1216.

Lew, A. (2003). *Variational time integrators in computational solid mechanics*. Ph.D thesis, California Institute of Technology, Pasadena, California, USA.

Lew, A., Marsden, J. E., Ortiz, M., & West, M. (2003). Asynchronous variational integrators. *Archive for Rational Mechanics and Analysis*, *2*, 85–146.

Lew, A., Marsden, J. E., Ortiz, M., & West, M. (2004). Variational time integrators. *International Journal for Numerical Methods in Engineering*, *60*, 153–212.

Leyendecker, S., Ober-Blöbaum, S., Marsden, J. E., & Ortiz, M. (2007). Discrete mechanics and optimal control for constrained multibody dynamics. In *Proceedings of the 6th International Conference on Multibody Systems, Nonlinear Dynamics, and Control, ASME* (pp. 1–10).

Leyendecker, S., Marsden, J. E., & Ortiz, M. (2008). Variational integrators for constrained dynamical systems. *ZAMM—Journal of Applied Mathematics and Mechanics*, *88*(9), 677–708.

Leyendecker, S., Hartmann, C., & Koch, M. (2012). Variational collision integrator for polymer chains. *Journal of Computational Physics*, *231*(10), 3896–3911.

Luo, M. Q., Liu, H., & Li, Y. M. (2013). Seismic wave modeling with implicit symplectic method based on spectral factorization on helix. *Chinese Journal of Geophysics*, *44*(3), 376–385.

Macchelli, A., Melchiorri, C., & Stramigioli, S. (2009). Port-based modeling and simulation of mechanical systems with rigid and flexible links. *IEEE Transactions on Robotics*, *25*(5), 1016–1029.

Maddocks, J. H. & Overton, M. L. (1995). Stability theory for dissipatively perturbed Hamiltonian systems. *Communications on pure and applied mathematics*, *XLVIII*, 583–610.

Maeda, S. (1980). Canonical structure and symmetries for discrete systems. *Mathematica Japonica*, *25*, 405–420.

Maeda, S. (1982). Lagrangian formulation of discrete systems and concept of difference space. *Mathematica Japonica*, *27*, 345–356.

Mäkinen, J. (2004). *A Formulation for flexible multibody mechanics. Lagrangian geometrically exact beam elements using constrain manifold parametrization*. Ph.D thesis, Tampere University of Technology, Institute of Applied Mechanics and Optimization.

Mäkinen, J., & Marjamäki, H. (2005). Total lagrangian parametrization of rotation manifold. In *ENOC-2005, Fifth EUROMECH Nonlinear Dynamics Conference*, (pp. 522–530).

Manning, R. S., & Maddocks, J. H. (1999). Symmetry breaking and the twisted elastic ring. *Computer Methods in Applied Mechanics and Engineering*, *170*(3–4), 313–330.

Marsden, J. E. (1988). The hamiltonian formulation of classical field theory. *Contemporary Mathematics*, *71*, 221–235.

Marsden, J. E., & Hughes, T. J. R. (1983). *Mathematical foundations of elasticity*. Prentice-Hall.

Marsden, J. E., & Ratiu, T. (1999). *Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems*. Springer-Verlag Gmbh.

Marsden, J. E., & Wendlandt, J. M. (1997). chapter Mechanical systems with symmetry, variational principles, and integration algorithms. *Current and future directions in applied mathematics* (pp. 219–261). Boston: Birkhuser.

Marsden, J. E., & West, W. (2001). Discrete mechanics and variational integrators. *Acta Numerica*, *10*, 357–514.

Marsden, J. E., Patrick, G. W., & Shkoller, S. (1998). Multisymplectic geometry, variational integrators, and nonlinear pdes. *Communications in Mathematical Physics*, *199*(2), 351–395.

Marsden, J. E., Pekarsky, S., & Shkoller, S. (1999). Discrete euler-poincar and lie-poisson equations. *Nonlinearity*, *12*(6), 1647–1662.

Marsden, J. E., Pekarsky, S., & Shkoller, S. (2000). Symmetry reduction of discrete lagrangian mechanics on lie groups. *Journal of Geometry and Physics*, *36*(1–2), 140–151.

Marsden, J. E., Pekarsky, S., Shkoller, S., & West, M. (2001). Variational methods, multisymplectic geometry and continuum mechanics. *Journal of Geometry and Physics*, *38*, 253–284.

Mata, P. (2015). Explicit symplectic momentum-conserving time-stepping scheme for the dynamics of geometrically exact rods. *Finite Elements in Analysis and Design*, *96*, 11–22.

Mata, P., & Lew, A. (2011). Variational time integrators for finite dimensional thermo-elasto-dynamics without heat conduction. *International Journal for Numerical Methods in Engineering*, *88*(1), 1–30.

Mata, P., & Lew, A. (2012). Structure-preserving time integrators for thermo-elasticity with heat conduction. Abstract in the *European Congress on Computational Methods in Applied Sciences and Engineering* Vienna, Austria, 10–14 September.

Mata, P. & Lew, A. (2014). Variational integrators for the dynamics of thermo-elastic solids with finite speed thermal waves. *Journal of Computational Physics*, *257*(Part B), 1423–1443.

Mata, P., Oller, S., & Barbat, A. H. (2008). Dynamic analysis of beam structures considering geometric and constitutive nonlinearity. *Computer Methods in Applied Mechanics and Engineering*, *197*, 857–878.

Mata, P., Barbat, A. H., Oller, S., & Boroschek, R. (2009). Non-linear seismic analysis of rc structures with energy-dissipating devices. *International Journal for Numerical Methods in Engineering*, *78*(9), 1037–1075.

Maugin, G. A. (2000). Towards an analytical mechanics of dissipative materials. *Rendiconti del Seminario Matematico. Geometry, Continua and Microstuctures. Universita e Politecnico di Torino*, Torino, *58*(2), 171–180.

Maugin, G. A., & Kalpakides, V. K. (2002). A Hamiltonian formulation for elasticity and thermoelasticity. *Journal of Physics A: Mathematical and General*, *35*, 10775–10788.

McLachlan, R. I., Perlmutter, M., & Quispel, G. R. W. (2004). On the nonlinear stability of symplectic integrators. *BIT Numerical Mathematics*, *44*, 99–117.

Mcrobie, F. A., & Lasenby, J. (1999). Simo-Vu Quoc rods using Clifford algebra. *International Journal for Numerical Methods in Engineering*, *45*, 377–398.

Meyer, K. R., Hall, G. R., & Offin, D. (2009). *introduction to hamiltonian dynamical systems and the n-body problem* (2nd ed., Vol. 90)., Applied mathematical sciences. Springer.

Meyer, K. R., Palacián, J. F., & Yanguas, P. (2011). Geometric averaging of hamiltonian systems: Periodic solutions, stability, and KAM tori. *SIAM Journal on Applied Dynamical Systems (SIADS)*, *10*(3), 817–856.

Milenkovic, V. J., & Milenkovic, V. (1997). Rational orthogonal approximations to orthogonal matrices. *Computational Geometry: Theory and Applications*, *7*, 25–32.

Mishchenko, A., & Fomenko, A. (1988). *A course of differential geometry and topology*. Moscow: Mir Publisher.

Moser, J., & Veselov, A. P. (1991). Discrete versions of some classical integrable systems and factorization of matrix polynomials. *Communications in Mathematical Physics*, *139*(2), 217–243.

Nichols, K., & Murphey, T. D. (2008). Variational integrators for constrained cables. In *IEEE International Conference on Automation Science and Engineering, 2008. CASE 2008*, (pp. 802–807), Arlington, VA, August 2008.

Noether, E. (1918). Invariante variationsprobleme. *Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen, Math-phys. Klasse*, *VI.6*, 235–257.

Ober-Blöbauma, S., Tao, M., Cheng, M., Owhadi, H., & Marsden, J. E. (2013). Variational integrators for electric circuits. *Journal of Computational Physics*, *242*, 498–530.

Patrick, G. W., & Cuell, C. (2009). Error analysis of variational integrators of unconstrained lagrangian systems. *Numerische Mathematik*, *113*(2), 243–264.

Pavlov, D. (2009). *Structure-preserving discretization of incompressible fluids*. Ph.D thesis, California Institute of Technology, Pasadena, California.

Pavlov, D., Mullen, P., Tong, Y., Kanso, E., Marsden, J. E., & Desbrun, M. (2011). Structure-preserving discretization of incompressible fluids. *Physica D: Nonlinear Phenomena*, *240*(6), 443–458.

Pérez-Morán, A. (2005). *Formulaciones tangente y secante en análisis no lineal de vigas de Cosserat*. Ph.D thesis, Universitat Politècnica de Catalunya, Spain.

Poincaré, H. (1899). *Les Méthodes Nouvelles de la Mécanique Céleste. Tome III*. Gauthiers-Villars.

Romero, I. (2009). Thermodynamically consistent time-stepping algorithms for non-linear thermomechanical systems. *International Journal for Numerical Methods in Engineering*, *79*, 706–732.

Romero, I. (2010). Algorithms for coupled problems that preserve symmetries and the laws of thermodynamics part I: Monolithic integrators and their application to finite strain thermoelasticity. *Computer Methods in Applied Mechanics and Engineering*, *199*, 1841–1858.

Romero, I., & Armero, F. (2002). An objective finite element approximation of the kinematics of geometrically exact rods and its use in the formulation of an energymomentum conserving scheme in dynamics. *International Journal for Numerical Methods in Engineering*, *54*, 1683–1716.

Ryckman, R. & Lew, A. (2010). Explicit asynchronous contact algorithm for elastic-rigid body interaction. In *Proceedings of the First International Conference in Computational Contact Mechanics*.

Ryckman, R., & Lew, A. (2011). *Trends in computational contact mechanics*, chapter Explicit asynchronous contact algorithm for elastic-rigid body interaction. (Vol. 58, pp. 169–191)., Lecture notes in applied and computational mechanics. Berlin: Springer.

Ryckman, R. A., & Lew, A. J. (2012). An explicit asynchronous contact algorithm for elastic bodyrigid wall interaction. *International Journal for Numerical Methods in Engineering*, *89*(7), 869–896.

Sansour, C., & Wagner, W. (2003). Multiplicative updating of the rotation tensor in the finite element analysis of rods and shells—a path independent approach. *Computational Mechanics*, *31*(1–2), 153–162.

Schmidt, B., Leyendecker, S., & Ortiz, M. (2009). $\gamma-$convergence of variational integrators for constrained systems. *Journal of Nonlinear Science*, *19*, 153–177.

Simo, J. C. (1985). A finite strain beam formulation. the three-dimensional dynamic problem. part i. *Computer Methods in Applied Mechanics and Engineering*, *49*, 55–70.

Simo, J. C., & Miehe, C. (1992). Associative coupled thermoplasticity at finite strains: Formulation, numerical analysis and implementation. *Computer Methods in Applied Mechanics and Engineering*, *98*(1), 41–104.

Simo, J. C., & Tarnow, N. (1994). A new energy and momentum conserving algorithm for the nonlinear dynamics of shells. *International Journal for Numerical Methods in Engineering*, *37*(15), 2527–2549.

Simo, J. C., & Vu-Quoc, L. (1986). A three-dimensional finite-strain rod model. part II: Computational aspects. *Computer Methods in Applied Mechanics and Engineering*, *58*, 79–116.

Simo, J. C., & Vu-Quoc, L. (1988). On the dynamics in space of rods undergoing large motions—a geometrically exact approach. *Computer Methods in Applied Mechanics and Engineering*, *66*, 125–161.

Simo, J. C., & Wong, K. K. (1991). Unconditionally stable algorithms for rigid body dynamics that exactly preserve energy and momentum. *International Journal for Numerical Methods in Engineering*, *31*(1), 19–52.

Simo, J. C., Marsden, J. E., & Krishnaprasad, P. S. (1988). The hamiltonian structure of nonlinear elasticity: The material and convective representations of solids, rods, and plates. *Archive for Rational Mechanics and Analysis*, *104*(2), 125–183.

Simo, J. C., Tarnow, N., & Wong, K. K. (1992). Exact energy-momentum conserving algorithms and symplectic schemes for nonlinear dynamics. *Computer Methods in Applied Mechanics and Engineering*, *100*, 63–116.

Simo, J. C., Tarnow, N., & Doblare, M. (1995). Non-linear dynamics of three-dimensional rods: Exact energy and momentum conserving algorithms. *International Journal for Numerical Methods in Engineering*, *38*(9), 1431–1473.

Spurrier, R. A. (1978). Comment on singularity-free extraction of a quaternion from a direction-cosine matrix. *Journal of Spacecraft and Rockets*, *15*(4), 255–255.

Stavros, F. (2014). *Nonlinear Hamiltonian mechanics applied to molecular dynamics. Theory and computational methods for understanding molecular spectroscopy and chemical reactions*. Springer.

Stern, A., & Grinspun, E. (2009). Implicit-explicit variational integration of highly oscillatory problems. *Multiscale Modelling and Simulation*, *7*, 1779–1794.

Stoffer, D. (1997). On the qualitative behaviour of symplectic integrators part I: Perturbed linear systems. *Numerische Mathematik*, *77*(4), 535–547.

Stoffer, D. (1998). On the qualitative behaviour of symplectic integrators. part III. Perturbed integrable systems. *Journal of Mathematical Analysis and Applications*, *217*(2), 521–545.

Stuelpnagel, J. (1964). On the parametrization of the three-dimensional rotation group. *SIAM Review*, *6*, 422–430.

Tao, M., Owhadi, H., & Marsden, J. E. (2010). Nonintrusive and structure preserving multiscale integration of stiff ODEs, SDEs, and hamiltonian systems with hidden slow dynamics via flow averaging. *Multiscale Modeling and Simulation*, *8*(4), 1269–1324.

Van Bargena, H., & Dimitroff, G. (2009). Isotropic ornstein-uhlenbeck flows. *Stochastic Processes and their Applications*, *119*(7), 2166–2197.

Veselov, A. P. (1988). Integrable discrete-time systems and difference operators. *Functional Analysis and Its Applications*, *22*(2), 83–93.

Vujanovic, B., & Djukic, D. J. (1971). On the variational principle of Hamilton's type for nonlinear heat transfer problem. *International Journal of Heat Mass Transfer*, *15*, 1111–1123.

Wang, L. (2007). *Variational integrators and generating functions for stochastic Hamiltonian systems*. Ph.D thesis, Universität Karlsruhe, Germany.

Wang, L., Hong, J., Scherer, R., & Bai, F. (2009). Dynamics and variational integrators of stochastic Hamiltonian systems. *International Journal of Numerical Analysis and Modeling*, *6*(4), 586–602.

Wendlandt, J. M., & Marsden, J. E. (1997). Mechanical integrators derived from a discrete variational principle. *Physica D: Nonlinear Phenomena*, *106*(3–4), 223–246.

West, W. (2004). *Variational integrators*. Ph.D thesis, California Institute of Technology, Pasadena, California, USA.

Wolff, S., & Bucher, C. (2013). Asynchronous collision integrators: Explicit treatment of unilateral contact with friction and nodal restraints. *International Journal for Numerical Methods in Engineering*, *95*(7), 562–586.

Yang, Q., Stainier, L., & Ortiz, M. (2006). A variational formulation of the coupled thermomechanical boundary-value problem for general dissipative solids. *Journal of the Mechanics and Physics of Solids*, *56*, 401–424.

Yoshida, H. (1993). Recent progress in the theory and application of symplectic integrators. *Celestial Mechanics and Dynamical Astronomy*, *56*(1–2), 27–43.