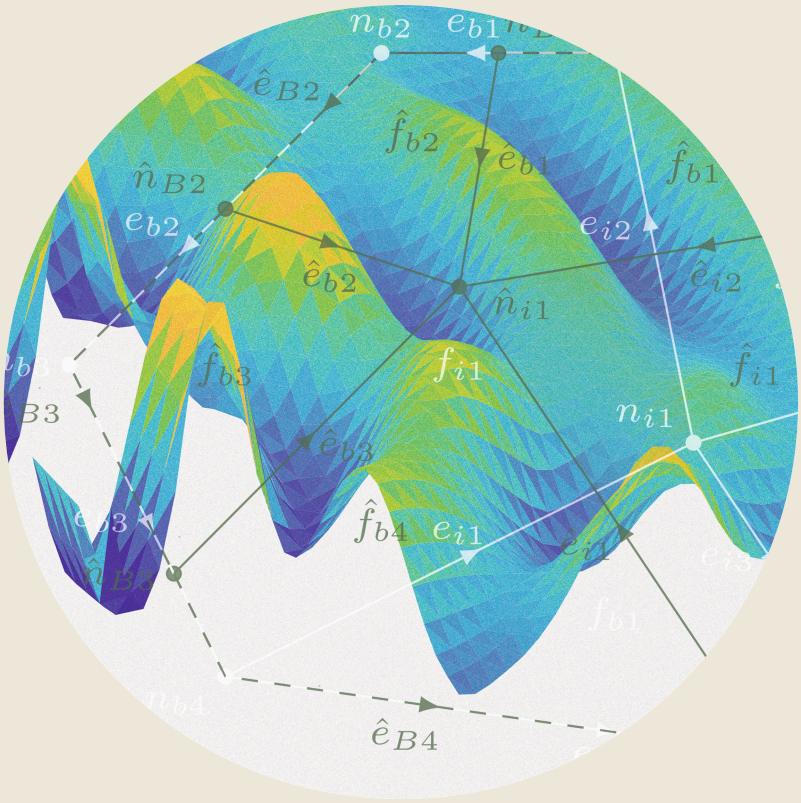

Paul Kotyczka

Numerical Methods for Distributed Parameter Port-Hamiltonian Systems

Maschinenbau

TUM.University Press

Paul Kotyczka

Numerical Methods for Distributed Parameter Port-Hamiltonian Systems

Structure-Preserving Approaches for Simulation and Control

The German National Library has registered this publication in the German National Bibliography. Detailed bibliographic data are available on the Internet at https://portal.dnb.de.

Imprint

Layout design and typesetting: Paul Kotyczka
Layout Guidelines for cover design: Designbuero Josef Grillmeier, Munich
Cover design: Caroline Ennemoser
Cover illustration: Paul Kotyczka, Caroline Ennemoser

TUM.University Press
Technical University of Munich
Arcisstrasse 21
80333 Munich

ISBN: 978-3-95884-028-7
DOI: 10.14459/2019md1510230

www.tum.de

To my family

Preface

The development of the *port-Hamiltonian* (PH) approach in the past three decades led to a structured, energy-based framework for modeling, system-theoretic analysis and control of interconnected multi-physics systems, which gathers researchers and practitioners from mathematics and engineering. With the growing complexity of the considered application cases – dimensionality, spatial extent, nonlinearity, non-trivial geometries and interconnections, or network aspects to mention only a few – also the use of numerical methods gains importance. In order to obtain consistently discretized models in space and/or time, which feature the characteristic *structural* properties of the original system, the numerical methods must be *structure-preserving*. In the context of PH systems, this means the conservation of a structural balance equation – in general for power – which also preserves the *modularity*, i. e. the compositionality via power ports, of the considered models.

This monograph, which is a slightly edited version of the submitted habilitation thesis, presents new approaches for the structure-preserving spatial discretization and numerical integration of PH systems. In addition to classical numerical techniques, the *open* character of PH systems must be taken into account. Pairs of in- and output variables, whose product represents the power flow through the system boundary must retain this interpretation in the finite-dimensional approximation. The vehicle to achieve structure-preservation is to maintain the separation of (i) the power interconnection structure, (ii) dynamics and (iii) the constitutive equations in the numerical scheme. This imposes, for example, a *mixed* approximation of the power variables or the discretization over *dual* meshes.

The contributions of this book extend the state of the art in several directions. Direct discrete modeling of systems of conservation laws on dual complexes is provided with the possibility to consider a *non-uniform* distribution of different (Dirichlet or Neumann) input boundary conditions (Chapter 3). The *weak form* of the Stokes-Dirac structure is defined as the basis for a mixed Galerkin approach, which yields finite-dimensional PH approximate models in an *explicit* state representation. Power-preserving mappings guarantee that the subspace of discrete port variables is a Dirac structure, i. e. is endowed with a non-degenerate power balance. The adequate definition of these mappings allows for a structure-preserving discretization of distributed parameter

PH systems in arbitrary spatial dimension, which is adapted to the nature of the system, in particular to its hyperbolic or parabolic character (Chapter 4). A new definition of *discrete-time* PH systems is introduced, which is based on the approximation of the power balance with the collocation method. The new definition generalizes existing approaches towards multi-stage schemes and makes the link to symplectic numerical integration methods for Hamiltonian systems (Chapter 5). Finally, the resulting numerical models are analyzed for the conservation of flatness of given outputs. It is shown on the examples of the 1D heat equation and 1D (non-)linear hyperbolic systems that the considered discretized models (in space and time) can be exploited for the computation of state and input trajectories with *explicit* numerical schemes (Chapter 6). In the whole book, the numerical quality of the approximate models is analyzed both by computation of consistency errors and/or numerical experiments.

The book summarizes my research between 2015 and 2018. I am indebted to Bernhard Maschke, who hosted me at the Laboratory of Automation and Process Engineering (LAGEP) in Lyon. Not only our excellent scientific exchange and the enthusiastic work on joint French-German projects made my research stay in France unique. He took also care of the well-being of my family and presented myself to *tout le monde*, which helped me quickly integrate in the French PH network. I thank my colleagues at LAGEP, in particular Boussad Hamroun, Françoise Couenne and Isabelle Pitault for their warm welcome and our discussions. I am grateful for the cordial collaboration with Laurent Lefèvre, who hosted me in Valence. Our whiteboard discussions were an important pillar of the work on numerical methods. *Merci* also to my friends and colleagues Yann Le Gorrec, Hector Ramírez, Denis Matignon and Flávio Cardoso for their exceptional hospitality during my visits in Besançon and Toulouse and the open spirit in our scientific exchange. I want to thank Prof. Boris Lohmann for his continuous support and the excellent working conditions at his Chair of Automatic Control at TUM. I thank my colleagues Mei Wang, Tobias Scheuermann and Hanae Labriji for their enthusiasm about our common research, as well as Regine Markwort, Ralf Hübner and Thomas Huber for our outstanding and friendly cooperation. Finally, I express my sincere gratitude to Prof. Andreas Kugi and Prof. Peter Eberhard for their valuable feedback and their suggestions, which had important influence on my work.

I gratefully acknowledge the financial support of the European Commission, Grenoble INP and Agence Nationale de la Recherche/Deutsche Forschungsgemeinschaft to enable our two years in France, which remain an unforgettable experience for our whole family.

I thank my parents for their help over all the years. I am proud of our children Katharina, Johannes and Magdalena, who make me the happiest father. Thank you, dear Christine, for your patience, your encouragement and your powerful support in every respect!

Olching, September 2019 *Paul Kotyczka*

Contents

Chapter 1

Introduction

1.1 Port-Hamiltonian Modeling and Control

The *port-Hamiltonian* (PH) approach has become a powerful framework for the *modeling, simulation and control of heterogeneous multi-physics systems*. The following paragraphs give an overview of the core concepts and highlight the main directions in theory and applications, which have been developed during the past three decades.

Modeling. *Port-Hamiltonian* systems' theory originates from the idea to extend the Hamiltonian formalism of analytical mechanics to network models of multi-domain physical systems, see [131] and the references therein, and to include *control* via *ports*, i. e. pairs of inputs and power-conjugated outputs [132], [193]. This unifying approach is intrinsically connected with the *bond graph* representation [154], [25], [66] of physical systems, which reveals the power flows (expressed through pairs of dual *port* variables, so-called *flows* and *efforts*) between subsystems and the environment, as well as the storage, conversion and dissipation of energy. The PH approach is therefore *modular* and the separation of the power-conserving interconnection *structure* from *dynamics* (energy storage), *dissipation* and the *constitutive equations* (which contain for example material laws or geometry parameters) is a key feature of the PH representation of dynamical systems. Interconnections and energy conversion are described mathematically by so-called *Dirac structures* [43], i. e. *linear* subspaces of power-conjugated port variables, on which power-conservation holds (for example Kirchhoff's laws in electrical networks or the combination of force balance and kinematic conditions for mechanical systems). The *Dirac structure* is the *underlying geometric structure* and therefore the "backbone" of a PH system. The *structural power balance* on a Dirac structure, in conjunction with dynamics and constitutive equations that are derived from the total energy – the *Hamiltonian* – implies *passivity* of a PH system. Linearity or *nonlinearity* of a finite-dimensional PH system with a constant interconnection structure depends on the Hamiltonian, which is quadratic in the former, and non-quadratic

in the latter case. Examples for finite-dimensional PH modeling of *open* physical systems are power systems networks [58], chemical reaction networks [194] or analog circuits for the simulation of electronic musical instruments [54], to mention only a few.

An important recent development is the formulation of *irreversible* PH systems [51], [162], [196], which allows to include the principles of *thermodynamics* in the structured framework for modeling and control. For irreversible processes, the *contact structure* relates the variables on the thermodynamic phase space.

The incorporation of *constraints*, which arise for example when numerical submodels of multi-physical systems are coupled, leads to *descriptor systems*, which are represented by *differential and algebraic equations* (DAEs). The PH formulation of DAE systems is an active field of current research, see e. g. [190], [13], [195].

The books [50], [191] and [192] give a large overview of the field of port-based modeling and passivity-based control for finite- and infinite-dimensional systems.

Control of finite-dimensional PH systems. Power ports as interfaces between system parts and the explicit use of *energy*[1] make the PH approach not only appealing for structured modeling and simulation of multi-physics, coupled systems, but also for control. The main mechanism behind *Control by Interconnection* (CbI) and *Interconnection and Damping Assignment Passivity-Based Control* (IDA-PBC) as introduced in [130], [150], [148] is *Energy Shaping*. In *CbI*, a system in PH representation is coupled with a PH dynamic controller in a *power-preserving* way, e. g. by a simple feedback interconnection. *Structural invariants* of the closed-loop systems, so-called *Casimir functions*, relate the states of plant and controller and are used to shape the (artificial) closed-loop energy function (the article [37] gives a characterization of the Casimirs under standard feedback interconnection). Lyapunov stability of the desired equilibrium is deduced from passivity of the closed loop and follows immediately if the shaped Hamiltonian has an isolated minimum in this equilibrium. Asymptotic stability can be achieved by passive output feedback (*damping injection*) under the assumption of zero state detectability [31].

Energy shaping may require a modification of the physical interconnection and damping structure (expressed in terms of a skew-symmetric and a symmetric positive semi-definite matrix, respectively), which leads to the IDA-PBC approach: Closed-loop desired target dynamics in PH form is matched with the open-loop system representation plus the unknown state feedback control law. The *matching conditions* restrict the assignable interconnection and damping matrices and the closed-loop Hamiltonian in such a way that the open-loop system must be *feedback-equivalent* [31] to a passive system with a suitable

[1] Called the *lingua franca* in [150], which *"facilitates communication [of practitioners] with control theorists"*.

positive-definite storage function. The survey articles [145] and [149] give an overview of CbI and IDA-PBC and the relations between both approaches. For linear PH system, the IDA-PBC matching conditions can be expressed in terms of LMIs [157]. [94] shows, how desired closed-loop *dynamics* can be incorporated in the IDA-PBC approach. The formulation of the mentioned energy- and passivity-based control techniques to *implicit* PH systems is the topic of [124]. [161] presents an IDA-PBC like approach for thermodynamic systems, where an *energy-based availability function* is shaped.

Passivity-based control techniques from the PH perspective find applications in electro-mechanical systems [12], power electronics and process control [85], to mention only a few examples. An important class of systems, to which energy shaping has been applied, are *underactuated mechanical systems* [147], [1], [203], [46]. The duality of PH passivity-based control with the technique of *Controlled Lagrangians* [17], [16], [146] is pointed out in [15].

Distributed parameter PH systems. The extension of port-Hamiltonian systems to the infinite-dimensional case was presented in [197] for *open* systems of two conservation laws, i. e. systems with energy flow through the boundary port. *Stokes-Dirac* structures[2] were introduced as Dirac structures on infinite-dimensional bond spaces. As in the finite-dimensional case, the Stokes-Dirac structure is the *linear infinite-dimensional space*, where the *pairs of power variables* on the spatial domain and on its boundary are related such that power conservation holds. The *in-* and *outputs* in the sense of systems' theory and control are defined via duality products whose values equal the exchanged powers at the ports. The constitutive equations, which distinguish between linear and nonlinear PH systems are derived from a Hamiltonian *functional* in the case of *hyperbolic* conservation laws. In [197], Maxwell's equations and the vibrating string are examples for systems that are described by a *canonical, formally skew-adjoint differential operator*, which represents the lossless conversion between energy forms. For flow problems, like the description of the ideal isentropic fluid or the rotational 2D shallow water equations, the Hamiltonian functional is non-quadratic and the canonical operator must be modified in order to account for the vorticity of the flow, see also [152].

Since the definition of distributed parameter port-Hamiltonian (dPH) systems in [197], a large number of works was dedicated to their mathematical and system theoretic analysis and the extension of energy-based control methods to *boundary control* of dPH systems in 1D. The article [112] deals with Dirac structures on (infinite-dimensional) Hilbert spaces based on generalized, higher order skew-symmetric differential operators. For a given choice of the differential operator, all possible pairs of boundary port variables are characterized to define an infinite-dimensional Dirac structure. Using this characterization, boundary conditions are given, such that the differential operator generates a contraction

[2] The prefix *Stokes* stems from the application of the generalized Stokes' theorem, which is instrumental in proving the structural properties of the linear subspace of power variables.

semigroup. Together with dynamics and linear constitutive equations (derived from a quadratic energy functional), dPH systems are defined in the sense of *boundary control systems* [44], Section 3.3, [57]. In [216], *well-posedness* and regularity of this class of dPH systems is shown under reasonable conditions on the system operator and the choice of boundary inputs[3]. The monograph [89] gives a concise introduction to linear dPH systems on Hilbert spaces over a one-dimensional domain with the semigroup approach. The thesis [201] exposes several aspects of mathematical modeling and properties of dPH systems, including the Riesz basis property for the 1D case, and presents prospective extensions of the semigroup approach to two- and three-dimensional domains. The recent article [106] treats the formulation of nD linear wave systems as boundary control systems. The definition of *distributed* power variables as in- and outputs is discussed in [140].

The port-based approach, introduced for the structured modeling of *open hyperbolic* systems, can also be applied to physical phenomena of different nature. *Parabolic* adsorption (diffusion) processes, as well as their couplings on different spatial scales, can be described based on the same Stokes-Dirac structures as a hyperbolic system of conservation laws [9]. The difference lies in the definition of dynamics and constitutive equations, which must account for the thermodynamic laws and properties. In [215], the authors show how to construct different systems, e. g. the parabolic heat equation, from the hyperbolic wave equation by using interconnections and closure equations. The main feature is that existence and uniqueness of the solution of the wave equation maps to existence and uniqueness of the resulting system. An impressive example for multi-physics modeling in the PH framework is the thermo-magneto-hydrodynamics model presented in [204] for the plasma in Tokamak fusion reactors. The PH formulation of the reactive Navier-Stokes flow [3] is based on a non-canonical and state-dependent skew-adjoint operator and the corresponding boundary port variables. A PH model of the compressible Euler equations in terms of density, weighted vorticity and dilatation is presented in [155]. Modeling of distributed parameter irreversible processes in the PH framework is the topic of [214]. In [141], "boundary multi-scale couplings" are introduced in the dPH model of a ionic-polymer-metal composite actuator in order to represent *unidirectional* energy flows at subsystem boundaries. [35] presents, as an example for fluid-structure interaction, the PH model of the sloshing liquid in a container coupled to a flexible beam.

Besides the PH formulation of conservation laws using Stokes-Dirac structures, this approach has also been applied to flexible mechanical systems, first for the Timoshenko beam model, which represents a hyperbolic system [126]. In [142], [129] and [189], nonlinear beam models, which are valid for large deflections, are presented based on Stokes-Dirac structures. An alternative approach for the PH representation of structure mechanical systems is the jet bundle

[3] In particular, not more than half of the boundary port variables can be imposed as inputs.

formulation [177], [172], which uses different state variables and is closer to the variational principles, from which the mechanical PDEs are derived [169]. In the Timoshenko beam example, the *configuration variables* deflection and rotation directly appear as states, as opposed to the *energy variables* shear and bending in the Stokes-Dirac approach [126]. For the PH modeling of the Mindlin plate (as the 2D correspondence of the Timoshenko beam) with both approaches, see [128] and [170].

Boundary control of dPH systems. The PH representation of beam models was accompanied by the design of boundary control in the sense of *Control by Interconnection* [126], [177], [171]. The infinite-dimensional system model is interconnected in a power-preserving way with a finite-dimensional controller in PH form, and the total energy functional is shaped using *Casimir functionals*, i. e. invariants of the mixed finite-infinite-dimensional closed-loop system. The opposite situation, namely the coupling of finite-dimensional PH systems over an infinite-dimensional interconnection structure is discussed in [127].

Damping injection can be realized via direct collocated boundary feedback of a passive output or suitable damping in the controller. The asymptotic convergence to a limit set is harder to prove in the infinite-dimensional setting than in finite dimensions. It requires to show precompactness of the orbit of the contraction semigroup generated by the system operator[4]. In [202], this argumentation is followed to prove asymptotic stability of a boundary control system connected to a static or strictly positive real dynamic controller. In [160], the approach is extended to exponential stability via the interconnection with a strictly input passive finite-dimensional controller. [136] shows the proof of asymptotic stability for a lossless Euler-Bernoulli beam under *nonlinear* dynamic boundary conditions. Another recent example with a rigorous proof of asymptotic stability is passivity-based damping control of a large scale multi-beam flexible manipulator is [80].

For finite-dimensional systems, the *dissipation obstacle* [150] states that a very simple class of passivity-based controllers (so-called *energy-balancing* controllers) can only be computed if the controller does not extract energy from the plant at the desired equilibrium. The dissipation obstacle is overcome by the use of state-modulated interconnections between plant and controller or the assignment of different interconnection and dissipation structures and the solution of a generalized matching equation (IDA-PBC). Recent approaches to translate these ideas to the boundary control of 1D dPH systems are described in [123], [125].

[4] See [119], Theorem 3.61, which leads to the infinite-dimensional formulation of LaSalle's invariance principle, Theorem 3.64.

1.2 Structure-Preserving Discretization

Along with the progress in theory and applications of PH systems, several directions have been explored to obtain finite-dimensional approximations of dPH systems under preservation of their structure. The same holds for the definition of discrete-time PH models, which are relevant for the high-fidelity simulation of PH systems and the implementation of passivity-based control in sampled control systems. We present an overview of the state of the art and comment on the relations to other *geometric* approaches.

Spatial discretization of dPH systems. The simulation and control by numerical methods, of *complex* (complex geometries, nonlinearities, interdomain couplings) distributed parameter PH systems requires a spatial discretization, which retains the underlying geometric properties related to *power continuity*. According to the separation of the interconnection structure from the dynamics and the constitutive equations, a *geometric* or *structure-preserving discretization* consists of three steps:

1. Finite-dimensional approximation of the underlying Stokes-Dirac structure. The duality between the *bond* or *power* variables (*flows* and *efforts*) must be mapped onto the finite-dimensional approximation. This requires a *mixed* approach with different approximation spaces for flows and efforts. The subspace of the approximated, discrete bond variables, on which the preserved *structural* continuity equation (for power in hyperbolic systems or entropy in irreversible systems) holds, defines a finite-dimensional *Dirac structure*.

2. Expression of the dynamics in terms of the discrete state variables.

3. Consistent discretization of the constitutive equations in the previously chosen approximation spaces. For hyperbolic systems, this gives rise to the definition of a *discrete Hamiltonian*, from which the discrete efforts (co-states) can be derived.

The first approach for a *structure-preserving* discretization of dPH systems in the spirit of *mixed finite elements* has been proposed in [71] for canonical hyperbolic systems. In [10], the approach was applied to a *diffusive* adsoption process. The mixed approximation bases lead to degeneracy of the discrete duality product (a bilinear form between flow and effort degrees of freedom), which is rectified by the definition of *reduced effort* variables. The Stokes-Dirac structure is discretized in *strong form*, which produces restrictive compatibility conditions. With the only admissible value of the discretization parameter[5], the resulting state space models feature dense matrices and a direct feedthrough, which is unnatural for hyperbolic (yet appropriate for parabolic) systems. In

[5] Using piecewise linear Whitney node forms for the efforts and piecewise constant edge forms for the flows, the only value for the parameter to define the reduced efforts according to [71], Eq. (18) such that the conditions of Proposition 1 are satisfied, is $\alpha_{ab} = \frac{1}{2}$.

the 1D *pseudo-spectral* method [138], Lagrange interpolation polynomials are used as basis functions. The approach is applied in [205] to generate a finite-dimensional approximate model of the radial resistive diffusion of the magnetic flux in Tokamak plasma. In this parabolic example, Bessel functions, which are eigenfunctions of a simplified problem, build the effort approximation basis.

In [56], [55], an alternative infinite-dimensional Dirac structure is defined to describe Maxwell's equations. In contrast to the canonical Stokes-Dirac structure, it contains the material parameters, i. e. it violates the strict separation of interconnection structure and constitutive equations. Compared to the dPH representation in [197], the role of state and co-state variable is permuted in one conservation law. The advantage is non-degeneracy of the discrete duality pairing, which retains its interpretation in terms of power if *compatible* approximation spaces (sequences of subspaces of the de Rham complex) are chosen. *Explicit* PH state space models are immediately obtained by invertibility of the corresponding finite element matrices.

The *weak* formulation as the basis for Galerkin numerical approximations, including the different variations of the finite element method (see [158], to cite only one textbook), has been only rarely used for modeling and discretization of PH systems: In [56], [55], one of the two conservation laws is written in weak form, including integration by parts. [3] presents the PH model of the reactive 1D Navier-Stokes equations in weak form. In [35], the inclusion of a piezo patch on a flexible beam in the PH model, and the structure-preserving discretization are performed via the weak form. The recent paper [34] presents a partitioned method, which is in the spirit of [56]. The application of integration by parts to only one of the two PDEs, which describe the 2D wave equation (in standard vector notation), and subsequent finite element discretization yields explicit state space models in PH form[6]. The recent articles [29], [30] deal with the PH modeling of plate models and the application of the partitioned finite element approach. An important aspect of using the weak form as basis for structure-preserving discretization is to make the link with well-known numerical methods and to pave the way for a simulation of PH systems with existing numerical tools like FreeFEM++ [79], GetDP [67] or FEniCS [2].

Other than the methods described above, the approach [175], [174] employs the language of *discrete exterior geometry* to formulate an integral representation of two conservation laws on a simplicial triangulation *and its dual*. Instead of using mixed or dual approximation spaces as in the finite element approach, the conservation laws are evaluated on *topologically dual objects*. The discrete linear constitutive relations are obtained by use of a *diagonal discrete Hodge operator*, which relates objects on both the primal and the dual mesh. The discrete topological objects (k-simplices and k-chains) on both meshes, as well as

[6] In [34], no compatibility conditions are considered for the approximation bases. Note however, that a compatible choice of finite element spaces, in accordance with the compatitibility condition in [56], and based on the geometric nature of the variables, seems, among other advantages, to avoid spurious modes (ongoing work).

linear functionals on them (e. g. the integrals of conserved quantities) are connected via boundary and coboundary mappings. The sequences of the discrete topological spaces, connected by the (co-)boundary maps represent discrete versions of the *de Rham* complex, so called n-complexes. [199] and [198] deal with the PH formulation of conservation laws on graphs and n-complexes. Related to this direct discrete geometric method are the recent structure-preserving approaches with finite volumes [95], [173] and finite differences [188] on *staggered grids*[7].

Spatial discretization of PDE systems yields finite-dimensional models of high order. *Structure-preserving model order reduction* generates low-order PH models for efficient simulation and control. Several approaches for linear PH systems have been presented in the last decade, e. g. [212], [156], [74], [69].

The use of finite-dimensional models from structure-preserving discretization for passivity-based feedback control is described for example in [122], [78] and [206]. In [99] and [207], the structure of the finite-dimensional approximations is used for inversion-based feedforward control design.

Relations with other geometric approaches. A *geometric* or structure-preserving discretization – as described for dPH systems above – is a *compatible* discretization as defined in [18]: *"Compatible discretizations transform partial differential equations to discrete algebraic problems that mimic fundamental properties of the continuum equations".* For dPH systems, such a fundamental property is the power balance, which holds on the Stokes-Dirac structure. Another fundamental property is the exact validity of the balance laws on discrete geometric objects. The *open* character of dPH systems requires special attention to the treatment of the boundary port variables. In particular, the *imposed* boundary conditions should appear as *inputs* in the resulting *control-oriented* models. This feature distinguishes structure-preserving approaches in the PH framework from other geometric discretization methods, some of which shall be mentioned here.

Bossavit's work in computational electromagnetism [20], [21] and Tonti's cell method [185] keep track of the geometric nature of the system variables which allows for a direct interpretation of the discrete variables in terms of *integral* system quantities. This integral point of view is also adopted in *discrete exterior calculus* [48]. *Finite element exterior calculus* [6] gives a theoretical frame to describe functional spaces of differential forms and their compatible approximations, which includes the construction of higher order approximation bases that generalize the famous *Whitney* forms [209], see also [163]. In the detailed survey on the finite element discretization of Maxwell's equations in frequency domain [83], the author points out that *"[t]o gain insight, a comprehensive view is mandatory, encompassing the structural aspects of the physical*

[7] See e. g. [153], Section 6.2 and 6.3 for the motivation to use staggered grids for the numerical solution of heat transfer and fluid flow problems.

model, a thorough knowledge of function spaces as well as familiarity with classical finite element techniques" (p. 238). Moreover, *"[f]inite elements that lack an interpretation as discrete differential forms*[8] *have to be used with great care"* (p. 240).

We refer also to the recent article [82] which proposes *conforming* polynomial approximation bases, in which the conservation laws are *exactly* satisfied, and which gives an excellent introduction to geometric discretization methods. In particular, an approximation of the field variables, which commutes with (exterior) differentiation *"guarantees that conservation and balance laws remain exactly satisfied in the discrete setting"* (p. 1456). Impressive examples for the use of geometric discretization methods can be found in weather prediction [41] or in the simulation of large-scale fluid flows [42], where the conservation of potential vorticity plays an important role.

Structure-preserving time integration. The *geometric* integration of ordinary differential equations, see e. g. [113], [76], is an important approach to perform long-time simulations of Hamiltonian systems. *Symplectic* integration conserves not only the symplectic form in the (mechanical) phase space, but also invariants of motion (Casimirs, first integrals). In [40], energy-preserving integrators for Hamiltonian systems with a non-canonical structure matrix (which can depend on the state) are introduced based on the collocation method. Symplectic integrators that are derived based on discrete versions of Hamilton's principle are called *variational integrators*. As indicated in the survey article [118], they *"work very well for both conservative and dissipative or forced mechanical systems"*. *Multi-symplectic* integrators are designed for the numerical solution of infinite-dimensional Hamiltonian systems, described by a multi-symplectic formulation of the underlying PDEs [28]. The *conservation law of symplecticity* of a hyperbolic system (like the shallow water equations) is preserved under appropriate numerical integration, e. g. the *Preissmann* box scheme [164].

For the structure-preserving numerical integration of PH systems, their *open* character has to be explicitly taken into account. This means that a discrete-time equivalent of the structural power balance must be found, which allows to approximate the energy transmitted over the input/output pair. The error of both this transmitted and the stored energy, is of fundamental interest in *interconnected* discrete-time PH systems, e. g. for multi-physics simulation.

Most existing works on the structure-preserving time integration or discrete-time formulation of PH systems make use of a *discrete gradient*, defined from a finite differences point of view [72], [4], [54]. A generic definition of PH dynamics on discrete manifolds (spaces that locally look like discretization grids or the set of floating-point numbers) is given in [182]. Objects and operations

[8] The discretization using Whitney elements features such an interpretation. The degrees of freedom can be considered to approximate the integrals of the conserved quantities over the finite integration domains.

from differential geometry are adapted to the discrete setting and discrete-time Dirac structures are defined. In the discrete setting, the chain rule is not valid, which means that the change of energy over a sampling interval is only *approximated* by a product of the discrete gradient and the increment of the state. The recent preprint [36] introduces higher order discrete gradient methods for explicit PH systems and introduces the notion of a discrete energy balance, which equates the approximations of energy supply and storage. However, no Runge-Kutta method satisfies such an *exact* discrete energy balance for arbitrary Hamiltonians.

Discrete-time passivity-based control. Some steps have been done towards the direct discrete-time design and implementation of passivity-based controllers in the PH framework. In [179], the passive interconnection between continuous-time and discrete-time PH systems is described in the context of robotics and telemanipulation, based on sampling and zero-order hold. The definition is based on the *exact* matching of exchanged energy per sampling interval. The implementation, however, relies on an approximation of the discrete-time effort variable in form of a delay. [108] presents a discrete-time IDA-PBC controller, which is implemented based on a forward Euler plant approximation and a discretization of the shaped potential energy gradient. [73] presents a comparable approach, based on the discrete gradient of the Hamiltonian. The paper [184] deals with the implementation of a continuous-time IDA-PBC controller in a sampled-data control system. The piecewise constant control input is computed based on the Taylor series expansion of the solution between two sampling instants.

1.3 Objectives of the Book

The following questions, which arise from the state of the art on structure-preserving discretization of PH systems, will be addressed in this book.

1. How can *different boundary causality* be invoked in a structured and versatile manner in the structure-preserving discretization of dPH systems? More precisely, how can finite-dimensional PH models in *explicit state space form* be obtained, which approximate systems of conservation laws under a *non-uniform* distribution of Dirichlet and Neumann boundary conditions? Until now, the direct discrete modeling approach on *dual chain complexes* [174], but also the finite-element method [56], [55] are formulated for a *uniform* type of imposed boundary conditions.

2. How can the *mixed finite-element* approach [71] be extended to higher spatial dimensions[9]? Moreover, the finite-dimensional approximation of a *hyperbolic* system of conservation laws (like the 1D telegraphers equations as

[9] Section 4 of [71] describes the *ansatz* to attack the 2D case on a triangular surface element. Working out the idea on a 2D simplicial mesh, however, yields a number of effort

a frequent example) with this approach, but also with the pseudo-spectral method [138], feature a *direct feedthrough* of the port variables at opposite boundaries. This effect is at odds with the finite propagation speed of information in a hyperbolic system and a questions is how to avoid it.

3. The finite element approach [71] is based on an approximation of the Stokes-Dirac structure in *strong form*, which leads to restrictive *compatibility conditions*, which leave no freedom to parametrize the approach. A *weak formulation* of the Stokes-Dirac structure relaxes these conditions. Being the basis for all Galerkin approximation methods, the weak form also allows to use more general approximation spaces, like the finite element spaces presented in [7]. Starting from a weak formulation therefore highlights and clarifies the relations of PH structure-preserving discretization with other *geometric* methods for the numerical approximation of PDEs.

4. The mixed finite element approach for nD dPH systems with non-uniform boundary causality, which is a key contribution of this monograph, uses *power-preserving mappings* of the discrete bond variables, which are equipped with parameters[10]. We analyze the effects of the parameter choice on the approximation quality of the finite-dimensional PH models in order to obtain suitable parametrizations for the structure-preserving discretization of both hyperbolic and parabolic systems.

5. For the simulation and computer-based control of finite-dimensional continuous-time models, time discretization is necessary. We explore how to use the ideas from spatial discretization – in particular the appropriate choice of bases for the bond variables – to structure-preserving time integration of PH systems. The objective is a clear definition of discrete-time PH systems, which (i) generalizes existing approaches in the PH framework, (ii) takes explicitly into account the numerical errors in the transmission and storage of energy and (iii) represents an extension of multi-stage symplectic integration schemes for Hamiltonian systems to the class of *open* PH systems.

6. The final objective of this monograph is to illustrate how to exploit the structure of the discretized PH models for numerical flatness-based feedforward control. We analyze conditions on the discretization parameters under which the flatness property of given outputs for the 1D heat and the 1D wave equation is preserved, and show the quality and convergence of the resulting feedforward controls in numerical experiments. For (nonlinear) hyperbolic systems, as opposed to the parabolic case, the combination of spatial *and* temporal discretization is necessary to obtain trajectories, which respect the unsmoothened transport of boundary and initial conditions.

---

degrees of freedom, which is *inferior* to the number of flow degrees of freedom. The original idea to map the *efforts* onto a space with identical dimension as the flow space does *not* work.

[10] In contrast to [71], our approach features tuning parameters in the *flow* mappings, which is the key ingredient for the extension to nD.

A natural question, which arises from the results presented in this monograph, is how to make use of the obtained models and the analysis of numerical errors for state observation and feedback control. This current field of research is, however, not treated in this book.

1.4 Outline

The book is structured as follows. **Chapter 2** is devoted to the basics of the PH formulation. We recall the Dirac structure and the Stokes-Dirac structure as backbones of the PH state representations of lumped and distributed parameter systems. For systems of two conservation laws – a canonical class of infinite-dimensional PH systems – we introduce a parametrization of the boundary in terms of the imposed (Dirichlet or Neumann) boundary conditions. The chapter closes with three characteristic examples, which can be represented in terms of the same Stokes-Dirac structure. We introduce the structured PH state representations of the nD linear wave equation, the 2D shallow water equations and the nD linear heat equation in both vector calculus notation and using differential forms.

In **Chapter 3**, we focus on the direct discrete representation of systems of conservation laws on dual, staggered meshes. After a section on the necessary preliminaries from discrete exterior calculus, we demonstrate how to construct the dual complexes (more precisely, the underlying staggered meshes) in order to incorporate a non-uniform distribution of input boundary conditions. The approach is presented in 2D, and the consistent numerical approximation of the constitutive equations is illustrated on the example of the irrotational shallow water equations. Remarks on the numerical approximation are given, which highlight the relations to existing works in the PH framework and to facts from numerics. The presented approach is currently being applied to heat transfer modeling in 3D catalytic foams.

The mixed Galerkin approximation of infinite-dimensional systems with a canonical differential operator is the topic of **Chapter 4**. Using the weak form of the Stokes-Dirac structure, the approximation problem is formulated in mixed bases. The degeneracy of the finite-dimensional power balance, which is expressed in terms of the flow and effort degrees of freedom, must be rectified with the definition of power-preserving mappings on the finite-dimensional bond space. The resulting Dirac structure and a consistent approximation of the constitutive equations are the ingredients for the finite-dimensional PH approximation in explicit state space form. The approach is illustrated with Whitney finite elements as approximation bases. For the 1D wave equation, the approximation quality of the obtained models is verified by a numerical convergence analysis of the eigenvalues and the comparison with the results of [71], both under variation of the design parameter. Accordingly, an optimal, centered parameter value is determined for the 1D heat equation, based on the analytic expressions for the eigenvalues of the finite-dimensional model. For the 2D wave equation, the construction of the power-preserving mappings and

the discretization of the constitutive equations are shown on a regular simplicial triangulation over a rectangular mesh. A simulation study gives evidence of the beneficial effect of *upwinding* numerical approximations for hyperbolic systems. Moreover, the interpretation of the discrete state variables, which is straightforward when using Whitney forms, constrains the 2D design parameters to reasonable values with regards to the approximation of vorticity. The simulation of the double slit experiment proves the applicability of the approach to non-trivial spatial domains.

Chapter 5 is dedicated to a new definition of discrete-time PH systems, which arises from the structure-preserving integration of explicit continuous-time PH systems. The definition follows the paradigm of separating the system's linear power structure from the possibly nonlinear constitutive equations and dynamics. The power balance of a lossless PH system, a structural property of its underlying Dirac structure, is mapped to a discrete-time energy balance per time step. Both the energy supplied via the control port and the energy routed to the storage elements, are approximated in the discretized setting. We give conditions on (i) the system and (ii) the discretization scheme, under which the continuous-time Dirac structure maps to a discrete-time Dirac structure. We define a discrete-time PH system as the completion of the discrete-time Dirac structure with constitutive equations and a numerical integration scheme. The analysis of the energy errors (supplied vs. stored energy) and the proof of their consistency under two classes of symplectic integration schemes is an important part of the chapter, and is illustrated by the numerical experiments, which support the order proofs.

Finally, **Chapter 6** shows the application of both structure-preserving discretization in space and time to the trajectory planning for parabolic and hyperbolic 1D boundary control systems. We show that under appropriate choices of the discretization parameter in the mixed Galerkin scheme, the flatness of given outputs of the 1D heat and the 1D wave equation is conserved. For the parabolic heat equation, a differential parametrization of the boundary input via the (opposite) flat boundary output and a finite number of time derivatives fits to the nature of the system and mimics the fact that trajectory planning using the infinite-dimensional model requires an infinitely smooth output trajectory. In contrast to that, smoothness of the trajectory is an unnatural constraint for the flat output of a hyperbolic system. Combination of the continuous-time PH approximate model with the symplectic Euler integration scheme yields a discrete-time model for the linear wave equation, which allows for a parametrization of states and input in terms of the flat output and its past and future values. This notion of flatness, which is natural for discrete-time and hyperbolic systems, holds also in the considered case of nonlinear conservation laws. A suitable discretization of the nonlinear constitutive equations allows for a stable, explicit numerical scheme for trajectory planning, which is illustrated on the example of the 1D shallow water equations.

Each chapter closes with a summary of the presented results, conclusions and an outlook to ongoing work and further research directions.

Chapter 2

Structured Representation of Conservation Laws

2.1 Finite-Dimensional Port-Hamiltonian Systems

Before introducing infinite-dimensional PH systems, we present the basic notions for the PH representation of finite-dimensional systems.

2.1.1 Dirac Structures

A Dirac structure, whose definition and characterization are summarized below, can be considered as "*the geometrical notion formalizing general power-conserving interconnections*" [197]. The following definition corresponds to [43], Definition 1.1.1.

Definition 2.1 (Dirac structure)**.** Given the finite-dimensional linear space F over $\mathbb{R}$ or another field and its dual $E = F^*$ with respect to the duality pairing[1] $\langle\cdot|\cdot\rangle : F \times E \to \mathbb{R}$. Define the symmetric bilinear form

$$\langle\langle(\mathbf{f}_1, \mathbf{e}_1), (\mathbf{f}_2, \mathbf{e}_2)\rangle\rangle := \frac{1}{2}\left(\langle \mathbf{e}_1|\mathbf{f}_2\rangle + \langle \mathbf{e}_2|\mathbf{f}_1\rangle\right), \quad (\mathbf{f}_i, \mathbf{e}_i) \in F \times E,\ i = 1, 2. \tag{2.1}$$

A *Dirac structure* is a linear subspace $D \subset F \times E$ which is *maximally isotropic* under $\langle\langle\cdot,\cdot\rangle\rangle$.

Equivalently, a Dirac structure can be characterized as the subspace $D \subset F \times E$ which equals its orthogonal complement with respect to $\langle\langle\cdot,\cdot\rangle\rangle$: $D = D^\perp$, see [197], Definition 2.1. D is isotropic under $\langle\langle\cdot,\cdot\rangle\rangle$, if $\langle\langle(\mathbf{f}_1, \mathbf{e}_1), (\mathbf{f}_2, \mathbf{e}_2)\rangle\rangle = 0$ for all $(\mathbf{f}_1, \mathbf{e}_1), (\mathbf{f}_2, \mathbf{e}_2) \in D$, from which $D \subset D^\perp$ follows. If, in addition, for

[1] The dual space $E = F^*$ contains all linear maps from F to $\mathbb{R}$; they can be written as duality pairings. If F – like in the finite-dimensional case – is endowed with an inner product structure, see [191], Remark 6.6.1, F and E are isomorphic and the duality pairing can be identified with an inner product.

every $(\mathbf{f}_1, \mathbf{e}_1) \in D$ there exists *no* $(\mathbf{f}_3, \mathbf{e}_3) \notin D$ such that $\langle\langle(\mathbf{f}_1, \mathbf{e}_1), (\mathbf{f}_3, \mathbf{e}_3)\rangle\rangle = 0$, then D is *maximally* isotropic, and also $D^\perp \subset D$ is true, which implies $D = D^\perp$. The isotropy condition implies that

$$\langle\langle(\mathbf{f}, \mathbf{e}), (\mathbf{f}, \mathbf{e})\rangle\rangle = \langle \mathbf{e} | \mathbf{f} \rangle = 0 \qquad \forall\, (\mathbf{f}, \mathbf{e}) \in F \times E. \tag{2.2}$$

For the space of conjugated power variables $F \times E \cong \mathbb{R}^n \times \mathbb{R}^n$, this is indeed a power balance equation. For more details and the different representations of finite-dimensional Dirac structures (in the PH context), we refer to [191], [197]. For Dirac structures defined on Hilbert spaces, and their composition, see e. g. Chapter 5 of [70] and [107].

The usefulness of the *symmetric bilinear form* in the definition of a Dirac structure can be illustrated with the following examples from electrical circuits (see [192], Subsection 2.2.2 and Section 2.1) and simple mechanical systems.

Example 2.1 (Electrical circuits)**.** Consider two electrical circuits with identical *topology*, but different network elements. $\mathbf{I}_1, \mathbf{I}_2 \in \mathbb{R}^m$ denote the vectors of currents through the m branches, and $\mathbf{V}_1, \mathbf{V}_2 \in \mathbb{R}^m$ contain the voltages across the branches. With $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2 \in \mathbb{R}^k$ the vectors of node potentials and $\mathbf{B} \in \mathbb{R}^{k \times m}$ the incidence matrix of the circuit graph, the vectors of currents and voltages, which satisfy Kirchhoff's current and voltage laws, can be written

$$\mathbf{B}\mathbf{I}_i = \mathbf{0}, \qquad \mathbf{V}_i = \mathbf{B}^T \boldsymbol{\phi}_i, \qquad i = 1, 2. \tag{2.3}$$

It is straightforward to verify that

$$\mathbf{I}_1^T \mathbf{V}_2 + \mathbf{I}_2^T \mathbf{V}_1 = 0 \tag{2.4}$$

holds, i. e. the space of voltages and currents that satisfy Kirchhoff's laws defines a finite-dimensional Dirac structure. For $\mathbf{I}_1 = \mathbf{I}_2 = \mathbf{I}$, $\mathbf{V}_1 = \mathbf{V}_2 = \mathbf{V}$, we obtain $\mathbf{I}^T \mathbf{V} = 0$, i. e. Tellegen's theorem [183]. Note, however, that by (2.3) also

$$\mathbf{I}_1^T \mathbf{V}_2 = 0, \qquad \mathbf{I}_2^T \mathbf{V}_1 = 0 \tag{2.5}$$

hold. The space of voltages and currents on the electrical circuit is a *separable* Dirac structure (see [192], Definition 2.2).

Example 2.2 (Mechanical oscillators)**.** Now consider the interconnection of a mass with a spring as depicted in Fig. 2.1 to produce an elementary mechanical oscillator. For two different pairs of masses m_1, m_2 and springs of stiffness k_1, k_2, the *interconnection* conditions (balance of forces, equality of velocities) read

$$\underbrace{\begin{bmatrix} -v_{s1} \\ -F_{m1} \end{bmatrix}}_{-\mathbf{f}_1} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{\mathbf{J}} \underbrace{\begin{bmatrix} F_{s1} \\ v_{m1} \end{bmatrix}}_{\mathbf{e}_1}, \qquad \underbrace{\begin{bmatrix} -v_{s2} \\ -F_{m2} \end{bmatrix}}_{-\mathbf{f}_2} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{\mathbf{J}} \underbrace{\begin{bmatrix} F_{s2} \\ v_{m2} \end{bmatrix}}_{\mathbf{e}_2}. \tag{2.6}$$

The power variables at mass and spring are assigned the roles of effort or flow, according to the model of a unified energy storing element, see also [192], Table

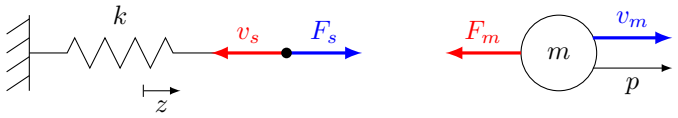

Figure 2.1: Spring and mass as lumped energy storage elements with flows (red) and efforts (blue).

B.1, and Fig. 2.2 in the following subsection. Other than for the electrical circuits, the mixed products of power variables are not zero in general,

$$\mathbf{e}_1^T \mathbf{f}_2 \neq 0, \qquad \mathbf{e}_2^T \mathbf{f}_1 \neq 0. \tag{2.7}$$

(To verify, consider forces and velocities in both oscillators as harmonic functions with different frequencies.) However, the *symmetrized* power product according to (2.1) (omitting the factor 1/2) equals zero:

$$\begin{aligned} \mathbf{e}_1^T \mathbf{f}_2 + \mathbf{e}_2^T \mathbf{f}_1 &= \mathbf{e}_1^T \mathbf{J} \mathbf{e}_2^T + \mathbf{e}_2^T \mathbf{J} \mathbf{e}_1 \\ &= \mathbf{e}_1^T (\mathbf{J} + \mathbf{J}^T) \mathbf{e}_2 \\ &= 0. \end{aligned} \tag{2.8}$$

For finite-dimensional Dirac structures, the proof that a subspace $D \subset F \times E$ is a Dirac structure is simplified, see [191], Proposition 6.6.4.

Theorem 2.1 (Finite-dimensional Dirac structure)**.** A subspace $D \subset F \times E$ on which $\langle \mathbf{e} | \mathbf{f} \rangle = 0$ holds for all $(\mathbf{f}, \mathbf{e}) \in D$ is a Dirac structure if and only if $\dim D = \dim F < \infty$.

In order to prove that a concrete subspace of power variables defines a finite-dimensional Dirac structure, the *kernel and image representation* are very useful, see [191], Proposition 6.6.6:

Theorem 2.2 (Kernel and image representation)**.** Given a Dirac structure $D \subset F \times E$ with $\dim F = n$. D admits

1. the *kernel representation*

$$D = \{(\mathbf{f}, \mathbf{e}) \in F \times E \mid \mathbf{F}\mathbf{f} + \mathbf{E}\mathbf{e} = \mathbf{0}\} \tag{2.9}$$

2. and the *image representation*

$$D = \{(\mathbf{f}, \mathbf{e}) \in F \times E \mid \exists \boldsymbol{\lambda} \in \mathbb{R}^n \text{ such that } \mathbf{f} = \mathbf{E}^T \boldsymbol{\lambda},\ \mathbf{e} = \mathbf{F}^T \boldsymbol{\lambda}\} \tag{2.10}$$

with matrices $\mathbf{E}, \mathbf{F} \in \mathbb{R}^{n \times n}$ that satisfy the following conditions:

(i) $\mathbf{E}\mathbf{F}^T + \mathbf{F}\mathbf{E}^T = \mathbf{0}$,

(ii) $\operatorname{rank} \begin{bmatrix} \mathbf{F} & \mathbf{E} \end{bmatrix} = n$.

Vice versa, the subspaces of $F \times E$ described by (2.9) and (2.10) are Dirac structures.

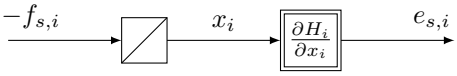

Figure 2.2: Variables in a canonical, independent energy storage element.

2.1.2 State Space Representation

If we include external port variables, i. e. inputs and power-conjugated outputs $\mathbf{u}, \mathbf{y} \in \mathbb{R}^m$, and associate internal flows and efforts to n separable energy storage elements according to the simple model depicted in Fig. 2.2, a kernel representation of the subspace (subscript "s" for "storage")

$$D = \{\mathbf{f}_s, \mathbf{e}_s \in \mathbb{R}^n, \mathbf{u}, \mathbf{y} \in \mathbb{R}^m \mid \mathbf{e}_s^T \mathbf{f}_s + \mathbf{u}^T \mathbf{y} = 0\} \tag{2.11}$$

is

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{f}_s \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} \mathbf{J} & \mathbf{G} \\ -\mathbf{G}^T & -\mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{e}_s \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{2.12}$$

with the skew-symmetric interconnection and feedthrough matrices $\mathbf{J} = -\mathbf{J}^T$, $\mathbf{D} = -\mathbf{D}^T$. Rearranging this system of equations, we obtain the *input-state-output representation* of the Dirac structure D

$$\begin{aligned} -\mathbf{f}_s &= \mathbf{J}\mathbf{e}_s + \mathbf{G}\mathbf{u} \\ \mathbf{y} &= \mathbf{G}^T\mathbf{e}_s + \mathbf{D}\mathbf{u}. \end{aligned} \qquad \text{(Structure)} \tag{2.13}$$

Adding both dynamics and constitutive equations

$$\dot{\mathbf{x}} = -\mathbf{f}_s \qquad \text{(Dynamics)} \tag{2.14a}$$

$$\mathbf{e}_s = \nabla H(\mathbf{x}) \qquad \text{(Constit. Eq.)} \tag{2.14b}$$

with $H(\mathbf{x})$ a *Hamiltonian function*[2] with strict minimum in $\mathbf{x}^* = \arg\min_{\mathbf{x}} H(\mathbf{x})$, we obtain the *explicit* state representation of a *port-Hamiltonian system* with constant interconnection, input and feedthrough matrices:

$$\dot{\mathbf{x}} = \mathbf{J}\nabla H(\mathbf{x}) + \mathbf{G}\mathbf{u} \tag{2.15a}$$

$$\mathbf{y} = \mathbf{G}^T \nabla H(\mathbf{x}) + \mathbf{D}\mathbf{u}. \tag{2.15b}$$

The differential energy balance

$$\begin{aligned} \dot{H} &= \frac{\partial H}{\partial \mathbf{x}} \dot{\mathbf{x}} \\ &= \mathbf{y}^T \mathbf{u} \end{aligned} \tag{2.16}$$

is a *structural* property, which can be directly obtained from $\mathbf{e}_s^T \mathbf{f}_s + \mathbf{u}^T \mathbf{y} = 0$ according to (2.11) and the substitution of dynamics and constitutive equations (2.14).

[2] In the case of n separable energy storage elements $H(\mathbf{x}) = \sum_{i=1}^{n} H_i(x_i)$.

Allowing for state-dependent matrices and introducing losses by a positive semi-definite dissipation matrix leads to the following definition[3] of *input-state-output PH systems*.

Definition 2.2 (PH system with dissipation)**.** A dynamical system of the form

$$\dot{\mathbf{x}} = (\mathbf{J}(\mathbf{x}) - \mathbf{R}(\mathbf{x}))\nabla H(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u} \tag{2.17a}$$

$$\mathbf{y} = \mathbf{G}^T(\mathbf{x})\nabla H(\mathbf{x}) + \mathbf{D}(\mathbf{x})\mathbf{u} \tag{2.17b}$$

with state vector $\mathbf{x} \in \mathbb{R}^n$, in- and outputs $\mathbf{u}, \mathbf{y} \in \mathbb{R}^m$, interconnection, damping and feedthrough matrices $\mathbf{J} = -\mathbf{J}^T$, $\mathbf{R} = \mathbf{R}^T \geq 0$ and $\mathbf{D} = -\mathbf{D}^T$ and an energy (Hamiltonian) function $H : \mathbb{R}^n \to \mathbb{R}$, which is bounded from below, is called an *input-state-output port-Hamiltonian* system.

From the structure of the equations and the definiteness properties, the power balance

$$\begin{aligned} \dot{H} &= -\frac{\partial H}{\partial \mathbf{x}}\mathbf{R}(\mathbf{x})\left(\frac{\partial H}{\partial \mathbf{x}}\right)^T + \mathbf{y}^T\mathbf{u} \\ &\leq \mathbf{y}^T\mathbf{u}, \end{aligned} \tag{2.18}$$

and hence *passivity*[4] of the PH state representation, immediately follows. If $\mathbf{x}^*$ is an *isolated* minimum of H, the Hamiltonian serves (at least locally) as a Lyapunov function for the stable equilibrium $\mathbf{x}^*$ of the unforced system. If the PH structure is imposed by control (e. g. using *Interconnection and Damping Assignment Passivity-Based Control*), the *shaped* Hamiltonian serves as closed-loop Lyapunov function.

Remark 2.1. A relevant class of PH systems are *implicit* systems, whose dynamics and constraints are given in terms of differential-algebraic equations (DAEs), see for example the articles [190] and [13]. PH DAE systems arise for example from automated network modeling of multi-physics systems. In this work, however, we focus on numerical methods that generate *explicit* state space approximate models for systems of conservation laws.

2.2 Systems of Conservation Laws

The dynamics of important classes of distributed parameter systems is governed by *conservation laws*. The describing equations are obtained by balancing the rate of change of *extensive quantities* (more precisely, their integrals), over finite

[3] See [191], Definition 6.6.1. Here, we add a simple feedthrough matrix $\mathbf{D}(\mathbf{x}) = -\mathbf{D}^T(\mathbf{x})$. For a more general definition of a PH system with feedthrough, see [191], Definition 6.6.2.

[4] See e. g. [191], Chapter 4.

spatial domains. The time derivative of these *conserved quantities*, which we consider as *state variables*, is induced by *fluxes* (again, we mean their integrals) over the boundary of the domain or in-domain *generation* terms. An example for the latter are reaction terms when balancing concentrations of chemical species. The *partial differential equations* (PDEs), which are a *local* description of the dynamics, result from the integrands that are left after the application of an *integral theorem* (the fundamental theorem of calculus, Stokes' theorem or Gauss' divergence theorem) to the boundary term[5]. This book is about *control-oriented* numerical methods for *open* systems of conservation laws, i. e. systems with a *boundary energy flow* that is imposed by *boundary control*. The port-Hamiltonian (PH) framework is particularly well-suited for open systems, and the presented methods for *spatial* and *temporal* discretization are derived in order to preserve the favorable PH structure of their mathematical description.

2.2.1 The Stokes-Dirac Structure

The PH formulation of a class of *open distributed parameter* systems was first introduced in [197], based on the definition of an infinite-dimensional Dirac structure. The *Stokes-Dirac structure* is a subspace of distributed power variables, which is maximally isotropic with respect to a symmetrized duality pairing of differential forms. Before we give a generalized definition of the Stokes-Dirac structure, which takes into account *non-uniform boundary causality*, we recall some definitions and the main theorem of [197].

We will use the notation in terms of *differential forms*, which is very natural for conservation laws[6]. A short summary of exterior calculus with differential forms, including references for further reading, is given in Section A.1 of the Appendix.

Consider an n-dimensional smooth manifold Ω with a smooth $(n-1)$-dimensional boundary $\partial\Omega$. The *natural* or *duality pairing* between two smooth differential forms[7] $\alpha \in \Lambda^k(\Omega)$ and $\beta \in \Lambda^{n-k}(\Omega)$ is given according to (A.3) by

$$\langle \beta | \alpha \rangle_\Omega := \int_\Omega \beta \wedge \alpha. \tag{2.19}$$

The pairing is *non-degenerate* as $\beta = 0$ ($\alpha = 0$) follows if $\langle \beta | \alpha \rangle_{\partial\Omega} = 0$ for all α (for all β). The duality pairing is defined accordingly on the boundary: $\langle \beta | \alpha \rangle_{\partial\Omega} := \int_{\partial\Omega} \beta \wedge \alpha$ for differential forms $\alpha \in \Lambda^k(\partial\Omega)$, $\beta \in \Lambda^{n-1-k}(\partial\Omega)$.

[5] This approach is in contrast to structural mechanics, where the PDEs are derived from variational principles.

[6] The reason is that conservation laws are defined via integrals over a balance region Ω and its boundary $\partial\Omega$. Differential k-forms can be clearly associated with integration over a k-dimensional domain. Moreover, the different integration theorems can be expressed in a unified manner with differential forms: the "Newton-Leibniz-Gauss-Green-Ostrogradskii-Stokes-Poincaré formula" ([8], §36.D) or simply the *generalized* Stokes' theorem.

[7] We will further specify the differentiability assumptions on the differential forms in the next subsection.

Define the spaces of *distributed flows and efforts*

$$\mathcal{F} := \Lambda^p(\Omega) \times \Lambda^q(\Omega) \times \Lambda^{n-p}(\partial\Omega), \tag{2.20a}$$
$$\mathcal{E} := \Lambda^{n-p}(\Omega) \times \Lambda^{n-q}(\Omega) \times \Lambda^{n-q}(\partial\Omega), \tag{2.20b}$$

where the scalars p and q satisfy

$$p + q = n + 1, \tag{2.21}$$

and the symmetrized duality pairing on $\mathcal{F} \times \mathcal{E}$

$$\langle\langle (f_1^p, f_1^q, f_1^\partial, e_1^p, e_1^q, e_1^\partial), (f_2^p, f_2^q, f_2^\partial, e_2^p, e_2^q, e_2^\partial)\rangle\rangle :=$$
$$\langle e_1^p | f_2^p\rangle_\Omega + \langle e_1^q | f_2^q\rangle_\Omega + \langle e_1^\partial | f_2^\partial\rangle_{\partial\Omega} + \langle e_2^p | f_1^p\rangle_\Omega + \langle e_2^q | f_1^q\rangle_\Omega + \langle e_2^\partial | f_1^\partial\rangle_{\partial\Omega}, \tag{2.22}$$

with $(i = 1, 2)$

$$f_i^p \in \Lambda^p(\Omega), \qquad f_i^q \in \Lambda^q(\Omega), \qquad f_i^\partial \in \Lambda^{n-p}(\partial\Omega), \tag{2.23a}$$
$$e_i^p \in \Lambda^{n-p}(\Omega), \qquad e_i^q \in \Lambda^{n-q}(\Omega), \qquad e_i^\partial \in \Lambda^{n-q}(\partial\Omega). \tag{2.23b}$$

The following can be proven[8].

Theorem 2.3 (Stokes-Dirac structure)**.** Given the infinite-dimensional linear spaces $\mathcal{F}$ and $\mathcal{E}$ as defined in (2.20) with p and q satisfying (2.21). Let flows and efforts be related by

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & (-1)^r \mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix} \tag{2.24a}$$
$$\begin{bmatrix} f^\partial \\ e^\partial \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & (-1)^p \end{bmatrix} \begin{bmatrix} e^p|_{\partial\Omega} \\ e^q|_{\partial\Omega} \end{bmatrix} \tag{2.24b}$$

with $r = pq + 1$, and where $|_{\partial\Omega}$ denotes the restriction of the given smooth differential forms to the boundary. Then the subspace

$$\mathcal{D} = \{(f^p, f^q, f^\partial, e^p, e^q, e^\partial) \in \mathcal{F} \times \mathcal{E} \mid (2.24) \text{ holds}\} \tag{2.25}$$

is a Dirac structure, i. e. $\mathcal{D} = \mathcal{D}^\perp$ with respect to the symmetrized duality pairing (2.22).

As the *generalized Stokes' theorem* (in terms of differential forms, see Appendix A.1) is used to prove that $\mathcal{D} \subset \mathcal{D}^\perp$ and $\mathcal{D}^\perp \subset \mathcal{D}$, from which $\mathcal{D} = \mathcal{D}^\perp$ follows, the so-defined Dirac structure is called *Stokes-Dirac structure*. Instrumental in the proof is the *formal skew-adjointness* of the matrix differential

[8] [197], Theorem 2.1.

operator[9] in (2.24a), which generalizes $\mathbf{J} = -\mathbf{J}^T$ in the representation (2.13) of the finite-dimensional Dirac structure.

An immediate implication is that flows and efforts, which belong to the Stokes-Dirac structure, satisfy the *structural power balance*

$$\langle e^p | f^p \rangle_\Omega + \langle e^q | f^q \rangle_\Omega + \langle e^\partial | f^\partial \rangle_{\partial\Omega} = 0. \tag{2.26}$$

Hyperbolic systems of *two conservation laws* can be represented by means of the above-defined Stokes-Dirac structure, for example Maxwell's equations, the vibrating string or – with a modification of the differential operator – an ideal isentropic fluid (all examples from [197]). Also diffusive phenomena, like adsorption processes [10], can be modeled with the help of the canonical differential operator in (2.24a). The Stokes-Dirac structure serves also to couple different physical phenomena (described by hyperbolic or parabolic PDEs) as presented in [204] on the complex example of thermo-magneto-hydrodynamics of plasmas in Tokamak fusion reactors.

The boundary term $\langle e^\partial | f^\partial \rangle_{\partial\Omega}$ in (2.26) pairs two power variables, one of which is considered as *control input* imposed on $\partial\Omega$. The other, *dual* variable plays the role of the *collocated* and *power-conjugated* output. The assignment of these roles to the boundary power variables is referred to as *causality of the boundary port*. For *boundary control* in the sense of [57], either e^∂ or f^∂ can be assigned the role of the (distributed) *boundary input*. In the rest of the work, we will use the following convention.

Definition 2.3 (Boundary inputs and collocated outputs)**.** Boundary efforts, denoted e^∂ play the role of boundary inputs, i. e. *imposed* boundary conditions. Boundary flows f^∂ are considered the power-conjugated, collocated outputs.

2.2.2 Non-Uniform Boundary Causality

To allow for a different causality along the boundary of the spatial domain, in other words to define different effort variables on the boundary as inputs, a generalized definition of the Stokes-Dirac structure has been given in [104]. We recall this definition, which departs from the specification of the functional spaces[10] of flows and efforts on the n-dimensional spatial domain Ω and its Lipschitz boundary $\partial\Omega$. With square integrability of the flow and effort differential

[9] A *formal* differential operator $\mathcal{J}$ is defined *without* boundary conditions (see e. g. [92], Sect. III.3). Formal skew-adjointness is verified by $\langle \mathbf{e} | \mathcal{J}\mathbf{e} \rangle_\Omega = -\langle \mathcal{J}\mathbf{e} | \mathbf{e} \rangle_\Omega$ under *zero* boundary conditions. The matrix operator in (2.24a) is formally skew-adjoint as for $e^p|_{\partial\Omega} = 0$ and $e^q|_{\partial\Omega} = 0$, we obtain (using integration by parts) $\langle e^p | (-1)^r \mathrm{d}e^q \rangle_\Omega + \langle e^q | \mathrm{d}e^p \rangle_\Omega = 0$.

[10] The functional spaces are particularly important for the compatible numerical approximation in Chapter 4.

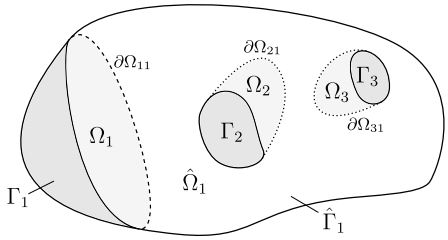

Figure 2.3: Sketch of a domain $\Omega \subset \mathbb{R}^3$ with subdomains $\Omega_1, \Omega_2, \Omega_3$ and $\hat{\Omega}_1$ and a partition of the boundary into $\Gamma_1, \Gamma_2, \Gamma_3$ and $\hat{\Gamma}_1$.

forms and weak differentiability of the latter[11]

$$\begin{aligned} f^p &\in L^2\Lambda^p(\Omega), & e^p &\in H^1\Lambda^{n-p}(\Omega), \\ f^q &\in L^2\Lambda^q(\Omega), & e^q &\in H^1\Lambda^{n-q}(\Omega), \end{aligned} \tag{2.27}$$

the boundedness of $\langle e^p | f^p \rangle_\Omega$ and $\langle e^q | f^q \rangle_\Omega$ is guaranteed, as well as the square integrability (in the Lebesgue sense) of $\mathrm{d}e^p$ and $\mathrm{d}e^q$. The *trace theorem* from classical functional analysis applies also to differential forms, see Appendix A.1 and references therein, which guarantees that the extensions to the boundary of the effort differential forms (denoted by the *trace* operator tr)

$$\operatorname{tr} e^p \in L^2\Lambda^{n-p}(\partial\Omega), \qquad \operatorname{tr} e^q \in L^2\Lambda^{n-p}(\partial\Omega) \tag{2.28}$$

are again square integrable. This ensures boundedness of $\langle \operatorname{tr} e^q | \operatorname{tr} e^p \rangle_{\partial\Omega}$.

Now consider an n-dimensional open and connected domain Ω with Lipschitz boundary $\partial\Omega$. Consider a partition of $\partial\Omega$ with subsets $\Gamma_i \subset \partial\Omega$, $i = 1, \ldots, n_\Gamma$, and $\hat{\Gamma}_j \subset \partial\Omega$, $j = 1, \ldots, \hat{n}_\Gamma$, with orientation according to $\partial\Omega$. Let $\bigcup_{i=1}^{n_\Gamma} \Gamma_i \cup \bigcup_{j=1}^{\hat{n}_\Gamma} \hat{\Gamma}_j = \partial\Omega$ and the intersections $\Gamma_i \cap \hat{\Gamma}_j$ be sets of measure zero. An illustration of such a domain Ω with the different portions of the boundary is given in Fig. 2.3. Define the boundary efforts (inputs) and flows

$$\begin{aligned} e_i^\Gamma &= (-1)^p \operatorname{tr} e^q|_{\Gamma_i}, & \hat{e}_j^\Gamma &= \operatorname{tr} e^p|_{\hat{\Gamma}_j}, \\ f_i^\Gamma &= \operatorname{tr} e^p|_{\Gamma_i}, & \hat{f}_j^\Gamma &= (-1)^p \operatorname{tr} e^q|_{\hat{\Gamma}_j}, \end{aligned} \tag{2.29}$$

as extensions of the effort forms to the corresponding subsets of $\partial\Omega$. The spaces of flows and efforts on Ω and its boundary (2.27) and (2.29) define the *bond space*[12], which is composed of

$$\begin{aligned} \mathcal{F} = L^2\Lambda^p(\Omega) \times L^2\Lambda^q(\Omega) \times L^2\Lambda^{n-p}(\Gamma_1) \times \cdots \times L^2\Lambda^{n-p}(\Gamma_{n_\Gamma}) \\ \times L^2\Lambda^{n-q}(\hat{\Gamma}_1) \times \cdots \times L^2\Lambda^{n-q}(\hat{\Gamma}_{\hat{n}_\Gamma}) \end{aligned} \tag{2.30}$$

[11] To be more precise, these properties refer to the coefficient functions of the differential forms, see [7], Section 1.

[12] As a reference to *bond graph modeling* of dynamical systems [154], see e. g. [50], Chapter 8.

and

$$\mathcal{E} = H^1\Lambda^{n-p}(\Omega) \times H^1\Lambda^{n-q}(\Omega) \times L^2\Lambda^{n-q}(\Gamma_1) \times \cdots \times L^2\Lambda^{n-q}(\Gamma_{n_\Gamma}) \\ \times L^2\Lambda^{n-p}(\hat{\Gamma}_1) \times \cdots \times L^2\Lambda^{n-p}(\hat{\Gamma}_{\hat{n}_\Gamma}). \quad (2.31)$$

The following theorem corresponds to Propostion 2.1 in [104].

Theorem 2.4 (Stokes-Dirac structure with non-uniform causality)**.** Consider the bond space $\mathcal{F} \times \mathcal{E}$ defined by (2.30) and (2.31) with in-domain flows and efforts (2.27) and boundary port variables (2.29). The subspace $\mathcal{D} \subset \mathcal{F} \times \mathcal{E}$, where

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & (-1)^r \mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix} \quad (2.32)$$

holds, is a Dirac structure. Flows and efforts satisfy the structural power balance

$$\langle e^p | f^p \rangle_\Omega + \langle e^q | f^q \rangle_\Omega + \sum_{i=1}^{n_\Gamma} \langle e_i^\Gamma | f_i^\Gamma \rangle_{\Gamma_i} + \sum_{j=1}^{\hat{n}_\Gamma} \langle \hat{f}_j^\Gamma | \hat{e}_j^\Gamma \rangle_{\hat{\Gamma}_j} = 0. \quad (2.33)$$

Proof. The proof follows the same lines as for [197], Theorem 2.1. It is shown both that $\mathcal{D} \subset \mathcal{D}^\perp$ and $\mathcal{D}^\perp \subset \mathcal{D}$, now with the modified symmetrized duality pairing $\langle\langle (\cdot,\cdot),(\cdot,\cdot) \rangle\rangle$, which is adapted to the different definition of $\mathcal{F} \times \mathcal{E}$, and taking into account that the power exchange over the boundary can be decomposed into the contributions of each subset $\Gamma_1, \ldots, \Gamma_{n_\Gamma}$ and $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{\hat{n}_\Gamma}$:

$$\sum_{i=1}^{n_\Gamma} \langle e_i^\Gamma | f_i^\Gamma \rangle_{\Gamma_i} + \sum_{j=1}^{\hat{n}_\Gamma} \langle \hat{f}_j^\Gamma | \hat{e}_j^\Gamma \rangle_{\hat{\Gamma}_j} = (-1)^p \langle \mathrm{tr}\, e^q | \mathrm{tr}\, e^p \rangle_{\partial\Omega}. \quad (2.34)$$

An alternative, as shown in the proof of [104], is to exploit the *compositionality* property of Stokes-Dirac structures, see also Remark 2.2 of [197]. □

Remark 2.2. In the above theorem, boundary efforts (i. e. the boundary inputs)

$$e_i^\Gamma =: u_i^q, \qquad \hat{e}_j^\Gamma =: u_j^p \quad (2.35)$$

and boundary flows (i. e. power-conjugated outputs)

$$f_i^\Gamma =: y_i^p, \qquad \hat{f}_j^\Gamma =: y_j^q \quad (2.36)$$

are defined as pure restrictions of either of the distributed efforts to the corresponding subsets Γ_i, $i = 1, \ldots n_\Gamma$ and $\hat{\Gamma}_j$, $j = 1, \ldots, \hat{n}_\Gamma$ of the boundary. It is, however, also possible to define images of the previous ones under a transformation that preserves the inner product (isometry), e. g. *scattering* variables [112].

2.2.3 Port-Hamiltonian Representation

We consider systems of two conservation laws in a *canonical form*[13] as introduced in [197]. These systems share a common *Stokes-Dirac structure*. From skew-adjointness of the canonical matrix differential operator as given in (2.24a), the *structural power balance* (2.26) or – with the definition of boundary port variables – (2.33) follows.

To define a port-Hamiltonian distributed parameter system, the Stokes-Dirac structure is completed by *dynamic equations* that introduce evolution with respect to time, and *constitutive relations*, which define the *nature* of the resulting dynamic system of PDEs. For *hyperbolic* system of conservation laws in PH form, the constitutive equations for the effort variables are derived from a single energy (Hamiltonian) functional.

The flows induce the time evolution of the distributed *state variables*[14] $p(z,t) \in L^2\Lambda^p(\Omega)$, $q(z,t) \in L^2\Lambda^q(\Omega)$ with corresponding initial conditions:

$$\begin{bmatrix} -\partial_t p(z,t) \\ -\partial_t q(z,t) \end{bmatrix} = \begin{bmatrix} f^p(z,t) \\ f^q(z,t) \end{bmatrix}, \qquad \begin{bmatrix} p(z,0) \\ q(z,0) \end{bmatrix} = \begin{bmatrix} p_0(z) \\ q_0(z) \end{bmatrix}. \tag{2.37}$$

The closure or *constitutive equations* relate the state and *co-state* (or co-energy or effort) variables according to

$$\begin{bmatrix} e^p(z,t) \\ e^q(z,t) \end{bmatrix} = \begin{bmatrix} \delta_p H(p(z,t),q(z,t)) \\ \delta_q H(p(z,t),q(z,t)) \end{bmatrix}, \tag{2.38}$$

where the right hand side contains the *variational derivatives* of the *Hamiltonian* or *energy* functional

$$H(p(z,t),q(z,t)) = \int_\Omega \mathcal{H}(p(z,t),q(z,t),z) \tag{2.39}$$

with the *Hamiltonian density* n-form $\mathcal{H}$. The variational derivatives are the unique differential $(n-p)$-form $\delta_p H$ and $(n-q)$-form $\delta_q H$ that satisfy[15]

$$H(p+\delta p, q+\delta q) = H(p,q) + \int_\Omega \delta_p H \wedge \delta p + \delta_q H \wedge \delta q \; + \; o(\delta p, \delta q). \tag{2.40}$$

The following definition generalizes [197], Definition 2.2, by including different causality of the boundary ports on subsets of $\partial\Omega$.

[13] Or systems of two conservation laws with *canonical interdomain coupling*.

[14] We use the same symbols for the state variables (as differential forms) and their degrees, which should not provoke any confusion. We start with explicitly indicating the arguments (z,t), for the Hamiltonian can depend on the spatial variables, as in the case of the shallow water equations with variable bed profile, see Subsection 2.3.2. In the sequel, we will frequently omit the arguments.

[15] See e. g. [50], p. 232.

Definition 2.4 (Canonical dPH system)**.** We call

$$\begin{bmatrix} -\partial_t p \\ -\partial_t q \end{bmatrix} = \begin{bmatrix} 0 & (-1)^r \mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} \delta_p H \\ \delta_q H \end{bmatrix}, \tag{2.41}$$

with the Hamiltonian functional H defined in (2.39) and the boundary port variables

$$\begin{aligned} e_i^\Gamma &= (-1)^p \operatorname{tr}(\delta_q H)_{\Gamma_i}, & \hat{e}_j^\Gamma &= \operatorname{tr}(\delta_p H)_{\hat{\Gamma}_j}, \\ f_i^\Gamma &= \operatorname{tr}(\delta_p H)_{\Gamma_i}, & \hat{f}_j^\Gamma &= (-1)^p \operatorname{tr}(\delta_q H)_{\hat{\Gamma}_j}, \end{aligned} \tag{2.42}$$

a (canonical) *distributed parameter port-Hamiltonian system* on the n-dimensional spatial manifold Ω.

Imposing the boundary efforts – in accordance with Definition 2.3 – as *control inputs* on subsets of $\partial\Omega$ (and understanding the boundary flows as *observation* or *output*), makes the system representation (2.41), (2.42) a *boundary control system* in the sense of [57]. For 1D linear PH systems with a *generalized* skew-adjoint system operator, [112] gives conditions on the assignment of boundary in- and outputs for the system operator to generate a contraction semigroup. The latter is instrumental to show *well-posedness* of a linear PH system, see [89]. Essentially, *at most half the number of boundary port variables* can be imposed as control inputs for a well-posed PH system in 1D.

Taking $\delta p = \partial_t p$, $\delta q = \partial_t q$ as variations in (2.40), and omitting the higher order terms, the time derivative of the energy functional (2.39) reads

$$\begin{aligned} \dot{H} &= \int_\Omega \delta_p H \wedge \partial_t p + \delta_q H \wedge \partial_t q \\ &= \langle \delta_p H | \partial_t p \rangle_\Omega + \langle \delta_q H | \partial_t q \rangle_\Omega. \end{aligned} \tag{2.43}$$

Replacing $\partial_t p$, $\partial_t q$ according to (2.41) and using the integration-by-parts formula (A.7) yields

$$\dot{H} = (-1)^p \langle \delta_q H | \delta_p H \rangle_{\partial\Omega}. \tag{2.44}$$

Equating the right hand sides of the last two equations gives, together with the definition of boundary port variables (2.42), the power balance

$$\underbrace{\langle \delta_p H | -\delta_t p \rangle_\Omega + \langle \delta_q H | -\delta_t q \rangle_\Omega}_{\substack{\text{power extracted from} \\ \text{distributed storage}}} + \underbrace{\sum_{i=1}^{n_\Gamma} \langle e_i^\partial | f_i^\partial \rangle_{\Gamma_i} + \sum_{j=1}^{\hat{n}_\Gamma} \langle \hat{f}_j^\partial | \hat{e}_j^\partial \rangle_{\hat{\Gamma}_j}}_{\substack{\text{power supplied} \\ \text{over the boundary}}} = 0. \tag{2.45}$$

It follows immediately from substitution of dynamics (2.37) and constitutive equations (2.38) in the *structural* balance equation (2.33). The balance equation for the Hamiltonian functional H shows – in the case of H being a positive

definite (or at least non-negative) *storage* functional – *passivity* of the infinite-dimensional PH state representation[16].

Remark 2.3. Defining the *flux functions* $\beta^p = (-1)^r \mathrm{d}e^q$ and $\beta^q = \mathrm{d}e^p$ as in Section 3.4, it is evident that (2.41), (2.42) represents a *hyperbolic* system of two conservation laws (see e. g. the books [114], [11]). Note that we explicitly defined boundary port variables whose pairing describes a power flow over the system boundary. We therefore deal with *open* systems of conservation laws.

Remark 2.4. For the *same* Stokes-Dirac structure, PDE systems of different nature are obtained when flows and efforts are defined based on different dynamics and closure equations. For a quadratic Hamiltonian density $\mathcal{H}$ in p and q, the resulting *hyperbolic* PH system is linear, otherwise nonlinear. The linear case is treated e. g. in [89], where $\mathcal{H}$ is bounded and non-negative, and H serves as the energy norm on the corresponding Hilbert space. For different definitions of flows and efforts, in particular if both efforts are not derived from the same functional, the resulting PDE system becomes *parabolic*, see e. g. [215], which allows to represent diffusive phenomena with the same Stokes-Dirac structure. Note also the heat conduction example in [50], Section 4.2.2, or [9], and Subsection 2.3.3 of this work.

Remark 2.5. The division of the system variables into *flows* (i. e. time derivatives of *states*) and *efforts* (or *co-states*) stems from the *duality* arizing from the variational formula (2.40), see also (2.43). It takes into account their different geometric definition, such as the degree of the differential forms. Tonti, for example, distinguishes between *configuration* and *source* variables [185], which are *states* and *efforts* in our language. His *energy variables* are products of these dual quantities, whereas in our context, we build the *duality products* between flows and efforts in order to compute *powers*. The space of dual power variables contains pairs of *in- and output variables* (denoted boundary efforts and flows), which describe the energy flow over the system boundary and make the PH representation inherently *control oriented*. Recall the central feature of PH modeling and control, which is the separation of the linear relations between the power variables – described by a (Stokes-)Dirac structure – from the constitutive and dynamics equations.

All numerical methods presented in this book aim at the preservation of this structural property under spatial and temporal discretization.

2.3 Examples

Having introduced the formal definitions of the Stokes-Dirac structure and canonical systems of two conservation laws in PH form, we present the three

[16] Passivity is defined in complete analogy to the finite-dimensional case, see e. g. [31], Definition 2.4.

example systems that serve throughout the book to illustrate the application of the developed structure-preserving numerical methods. Each example system is first expressed in classical vector calculus notation. Subsequently, its representation in terms of differential forms is given. In this section, whose focus is the representation of the governing PDEs in a structured form, we do not specify initial and boundary conditions. Boundary conditions will be explicitly treated in the subsequent chapters in the discretized setting for the considered numerical schemes.

2.3.1 Wave Equation

We consider the linear wave equation with constant speed of propagation $c > 0$ on an open, bounded domain $\Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$, with Lipschitz continuous boundary $\partial\Omega$ ($\Delta(\cdot) = \operatorname{div}\operatorname{grad}(\cdot)$ denotes the Laplace operator):

$$\partial_t^2 x(\mathbf{z}, t) = c^2 \Delta x(\mathbf{z}, t), \qquad \mathbf{z} \in \Omega. \tag{2.46}$$

2.3.1.1 Vector Calculus Notation

By definition of the scalar and vector valued state variables

$$u(\mathbf{z}, t) = -\partial_t x(\mathbf{z}, t) \in \mathbb{R}, \qquad \mathbf{v}(\mathbf{z}, t) = \nabla x(\mathbf{z}, t) \in \mathbb{R}^n, \tag{2.47}$$

Eq. (2.46) can be written as a set of two partial differential equations of order one in time. Throughout this book, we will consider a representation of these two equations that is split into three components:

$$\begin{bmatrix} f^u(\mathbf{z}, t) \\ \mathbf{f}^v(\mathbf{z}, t) \end{bmatrix} = \begin{bmatrix} 0 & \operatorname{div} \\ \operatorname{grad} & \mathbf{0} \end{bmatrix} \begin{bmatrix} e^u(\mathbf{z}, t) \\ \mathbf{e}^v(\mathbf{z}, t) \end{bmatrix}, \qquad \text{(Structure)} \tag{2.48a}$$

$$\begin{bmatrix} \partial_t u(\mathbf{z}, t) \\ \partial_t \mathbf{v}(\mathbf{z}, t) \end{bmatrix} = \begin{bmatrix} -f^u(\mathbf{z}, t) \\ -\mathbf{f}^v(\mathbf{z}, t) \end{bmatrix}, \qquad \text{(Dynamics)} \tag{2.48b}$$

$$\begin{bmatrix} e^u(\mathbf{z}, t) \\ \mathbf{e}^v(\mathbf{z}, t) \end{bmatrix} = \begin{bmatrix} \delta_u H(u(\mathbf{z}, t), \mathbf{v}(\mathbf{z}, t)) \\ \delta_v^T H(u(\mathbf{z}, t), \mathbf{v}(\mathbf{z}, t)) \end{bmatrix}. \qquad \text{(Constit. Eq.)} \tag{2.48c}$$

The *structure* equations represent a *linear* relation between the *dual power variables*[17] $f^u \in \mathbb{R}$, $\mathbf{f}^v \in \mathbb{R}^n$ called *flows* and $e^u \in \mathbb{R}$, $\mathbf{e}^v \in \mathbb{R}^n$ called *efforts*. The *dynamics* equations describe the evolution of the *state* variables (or *energy* variables) $u \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^n$, which in the linear case is induced exclusively by the flows. Finally, the *constitutive* equations close the system of equations: The efforts (or *co-energy* or *co-state* variables) are derived from an *energy* or *Hamiltonian* functional on Ω over the states

$$H(u, \mathbf{v}) = \int_\Omega \mathcal{H}(u, \mathbf{v}, \mathbf{z})\, d\mathbf{z} \tag{2.49}$$

[17] This meaning of the variables will become clear right below.

with *Hamiltonian density* $\mathcal{H} : \mathbb{R} \times \mathbb{R}^n \times \Omega \to \mathbb{R}$. The *variational derivatives* $\delta_u H \in \mathbb{R}$ and $\delta_v^T H = (\delta_v H)^T \in \mathbb{R}^n$ are defined by the first variation[18]

$$H(u + \delta u, \mathbf{v} + \delta\mathbf{v}) = H(u, \mathbf{v}) + \int_\Omega \delta_u H \, \delta u + \delta_v H \cdot \delta\mathbf{v} \, d\mathbf{z} + o(\delta u, \delta\mathbf{v}) \quad (2.50)$$

of H. For the linear wave equation (2.46),

$$\mathcal{H}(u, \mathbf{v}) = \frac{1}{2}u^2 + \frac{1}{2}c^2\mathbf{v} \cdot \mathbf{v} \quad (2.51)$$

is the quadratic Hamiltonian density. It is straightforward to verify that the variational derivatives of H coincide with the partial derivatives of $\mathcal{H}$, if the Hamiltonian density depends only on the states (and possibly $\mathbf{z}$) and *not* on their spatial derivatives:

$$\delta_u H(u, \mathbf{v}) = \partial_u \mathcal{H}(u, \mathbf{v}, \mathbf{z}), \qquad \delta_v H(u, \mathbf{v}) = \partial_v \mathcal{H}(u, \mathbf{v}, \mathbf{z}). \quad (2.52)$$

An immediate consequence is the differential energy balance

$$\begin{aligned} \dot{H}(u, \mathbf{v}) &= \int_\Omega \partial_u \mathcal{H} \, \partial_t u + \partial_v \mathcal{H} \cdot \partial_t \mathbf{v} \, d\mathbf{z} \\ &= -\int_\Omega e^u f^u + \mathbf{e}^v \cdot \mathbf{f}^v \, d\mathbf{z}, \end{aligned} \quad (2.53)$$

which confirms the meaning of flows and efforts as power variables. Substitution of the structure equation (2.48a) in the right hand side term, yields, under application of Gauss' divergence theorem,

$$\begin{aligned} -\int_\Omega e^u f^u + \mathbf{e}^v \cdot \mathbf{f}^v \, d\mathbf{z} &= -\int_\Omega e^u \operatorname{div} \mathbf{e}^v + \mathbf{e}^v \cdot \operatorname{grad} e^u \, d\mathbf{z} \\ &= -\int_\Omega \operatorname{div} (e^u \mathbf{e}^v) \, d\mathbf{z} \\ &= -\int_{\partial\Omega} e^u \mathbf{e}^v \cdot \mathbf{n} \, dS, \end{aligned} \quad (2.54)$$

with $\mathbf{n}$ the unit normal vector on the boundary $\partial\Omega$ with surface element dS. The *power balance equation*

$$(e^u, f^u)_\Omega + (\mathbf{e}^v, \mathbf{f}^v)_\Omega + (e^u, -\mathbf{e}^v \cdot \mathbf{n})_{\partial\Omega} = 0 \quad (2.55)$$

with $(\cdot, \cdot)_\Omega$ and $(\cdot, \cdot)_{\partial\Omega}$ standard L^2 inner products on Ω and its boundary is a *fundamental structural property* of the considered (lossless) PH systems. It stems *only* from the formal skew-adjointness of the matrix differential operator in the structure equation (2.59a) and is *independent of* the actual definition of the effort variables by the constitutive equations (2.59c).

[18] The dot "$\cdot$" denotes the usual scalar product and $o(\delta u, \delta\mathbf{v})$ contains terms of order greater than one in δu, δv_i, $i = 1, \ldots, n$.

2.3.1.2 Exterior Calculus Notation

In particular for systems of conservation laws, the representation in terms of *differential forms* is physically insightful. Applying the rules of exterior calculus, as summarized in Appendix A.1, define the flow and state differential forms

$$\begin{aligned} f^p &:= *f^u \in \Lambda^n(\Omega), & f^q &:= (\mathbf{f}^v)^\flat \in \Lambda^1(\Omega), \\ p &:= *u \in \Lambda^n(\Omega), & q &:= \mathbf{v}^\flat \in \Lambda^1(\Omega), \end{aligned} \tag{2.56}$$

as well as the effort differential forms

$$e^p := e^u \in \Lambda^0(\Omega), \qquad e^q := (-1)^{n-1} * (\mathbf{e}^v)^\flat \in \Lambda^{n-1}(\Omega). \tag{2.57}$$

The factor $(-1)^{n-1}$ ensures that e^q can be derived from the energy functional, which is expressed in terms of *duality products* of differential forms:

$$\begin{aligned} H(p,q) &= \frac{1}{2}(p,p)_\Omega + \frac{1}{2}c^2(q,q)_\Omega \\ &= \frac{1}{2}\langle p | *p \rangle_\Omega + \frac{1}{2}c^2 \langle q | *q \rangle_\Omega. \end{aligned} \tag{2.58}$$

Its variational derivatives $\delta_p H$ and $\delta_q H$ are defined by Eq. (2.40). Equations (2.48a)–(2.48c) become

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & (-1)^{n-1}\mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix}, \qquad \text{(Structure)} \tag{2.59a}$$

$$\begin{bmatrix} \partial_t p \\ \partial_t q \end{bmatrix} = \begin{bmatrix} -f^p \\ -f^q \end{bmatrix}, \qquad \text{(Dynamics)} \tag{2.59b}$$

$$\begin{bmatrix} e^p \\ e^q \end{bmatrix} = \begin{bmatrix} \delta_p H \\ \delta_q H \end{bmatrix}, \qquad \text{(Constit. Eq.)} \tag{2.59c}$$

with the *exterior derivative* d instead of the vector calculus differential operators div and grad.

Remark 2.6. Although throughout the book we consider differential forms on subspaces of the Euclidean space (the identity of L^2 inner product and the duality product in (2.58) gives evidence for this), note that the representation via differential forms is *a priori coordinate-free*. Coordinates come only into play when integrals are explicitly computed or with *metric-dependent* operations like the Hodge star.

From the structure equation (2.59a) we obtain, by application of the integration-by-parts formula (A.7), the balance equation

$$\langle e^p | f^p \rangle_\Omega + \langle e^q | f^q \rangle_\Omega + (-1)^n \langle \mathrm{tr}\, e^q | \mathrm{tr}\, e^p \rangle_{\partial\Omega} = 0. \tag{2.60}$$

Substituting (2.59b) and (2.59c) in the first two terms of (2.60), we note that these amount for the (negative) time derivative of the energy functional H:

$$\begin{aligned} \langle \delta_p H | -\partial_t p \rangle_\Omega + \langle \delta_q H | -\partial_t q \rangle_\Omega &= -\int_\Omega \partial_p \mathcal{H} \wedge \partial_t p + \partial_q \mathcal{H} \wedge \partial_t q \\ &= -\partial_t H. \end{aligned} \tag{2.61}$$

With

$$\partial_t H = (-1)^n \langle \operatorname{tr} e^q | \operatorname{tr} e^p \rangle_{\partial\Omega}, \tag{2.62}$$

we observe that the change of stored energy is induced by the duality product on the boundary of the effort variables, which amounts to the *power supplied over the boundary*. The extension of the efforts to the boundary, which in particular clarifies their functional spaces, is denoted by the trace operator tr.

Example 2.3 (Telegrapher's equations)**.** The simplest 1D example of a system of two conservation laws is an electric transmission line with the spatial coordinate $z \in \Omega = (0, L)$, see e. g. [71]. With $p(z) = \psi(z) \in \Lambda^1(\Omega)$, the magnetic flux density one-form, $q(z) \in \Lambda^1(\Omega)$, the electric charge density one-form, $l(z)dz, c(z)dz \in \Lambda^1(\Omega)$ the distributed inductance and capacitance per length ($l(z)$ and $c(z)$ are smooth functions and dz the basis one-form), the Hamiltonian density one-form is $\mathcal{H}(p,q) = \frac{1}{2}\left(p(z) \wedge *\frac{p(z)}{l(z)} + q(z) \wedge *\frac{q(z)}{c(z)}\right)$. The *Hodge star* operator $* : \Lambda^k(\Omega) \to \Lambda^{n-k}(\Omega)$ renders in the 1D case a one-form a zero-form and *vice versa*. The variational derivatives of the Hamiltonian $H = \int_0^L \mathcal{H}$ are the current and the voltage along the line, $e^p(z) = \delta_p H = \frac{*p(z)}{l(z)} = i(z) \in \Lambda^0(\Omega)$ and $e^q(z) = \delta_q H = \frac{*q(z)}{c(z)} = v(z) \in \Lambda^0(\Omega)$. Note that in the 1D case, the boundary $\partial\Omega$ consists of two disconnected points. With $\langle \operatorname{tr} e^q | \operatorname{tr} e^p \rangle_{\partial\Omega} = e^q(L)e^q(L) - e^q(0)e^p(0)$, the structural balance equation (2.60) reads

$$\langle e^p | f^p \rangle_\Omega + \langle e^q | f^q \rangle_\Omega + e^q(0)e^p(0) - e^q(L)e^p(L) = 0. \tag{2.63}$$

Substitution of dynamics $\dot{p}(z) = -f^p(z)$, $\dot{q}(z) = -f^q(z)$ and constitutive equations yields the differential energy balance

$$\partial_t H = i(0)v(0) - i(L)v(L). \tag{2.64}$$

2.3.2 2D Shallow Water Equations

The shallow water equations describe the two-dimensional flow of an inviscid fluid with relatively low depth ("shallow"), which permits the averaging of the horizontal components of the velocity field and the omission of the vertical velocity component.

2.3.2.1 Vector Calculus Notation

The two equations that describe the conservation of mass and momentum over an infinitesimal, fixed surface element[19] (we consider the fluid in a non-rotating system) can be written in vector calculus notation, with spatial coordinates $\mathbf{z} = \begin{bmatrix} x & y \end{bmatrix}^T$, see e. g. [42], [104],

$$\begin{bmatrix} -\partial_t h \\ -(\partial_t \mathbf{u} + \zeta \mathbf{u}^\perp) \end{bmatrix} = \begin{bmatrix} 0 & \operatorname{div} \\ \operatorname{grad} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2}\mathbf{u} \cdot \mathbf{u} + g(h+b) \\ h\mathbf{u} \end{bmatrix}, \tag{2.65}$$

[19] Which corresponds to the *Eulerian* representation of the fluid flow.

where h denotes the water level over the bed, b is the elevation of the bed profile, $\mathbf{u} = [u \quad v]^T$ the 2-dimensional velocity field, $h\mathbf{u}$ the specific discharge vector and g the gravitational acceleration.

$$\zeta = \nabla^\perp \cdot \mathbf{u} := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \left(\nabla \times \begin{bmatrix} u \\ v \\ w \end{bmatrix} \right) = \partial_x v - \partial_y u \tag{2.66}$$

is the *vorticity* of the flow (w is the neglected vertical component of the velocity vector field), and

$$\mathbf{u}^\perp = \begin{bmatrix} -v \\ u \end{bmatrix} \tag{2.67}$$

is defined via the x and y component of $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \times \begin{bmatrix} u & v & w \end{bmatrix}^T$.

Remark 2.7. The term $\zeta \mathbf{u}^\perp$, which represents the acceleration due to the rotation of the flow[20], can equivalently be expressed as $PV(h\mathbf{u})^\perp$ with $PV = \frac{\zeta}{h}$. The *potential vorticity* satisfies the balance equation $\partial_t PV + \mathbf{u} \cdot \nabla PV = 0$, i. e. it is advected with the fluid flow, see e. g. [5]. It plays an important role in the long-time numerical simulation of large scale flow problems, see e. g. [165].

With the total energy (per unit mass)

$$H = \int_\Omega \mathcal{H} \, d\mathbf{z} \qquad \text{with} \qquad \mathcal{H} = \frac{1}{2} h\mathbf{u} \cdot \mathbf{u} + \frac{1}{2} g h^2 + ghb, \tag{2.68}$$

the shallow water equations can be rewritten in the structured form

$$\begin{bmatrix} f^h \\ \mathbf{f}^u \end{bmatrix} = \begin{bmatrix} 0 & \text{div} \\ \text{grad} & \mathbf{0} \end{bmatrix} \begin{bmatrix} e^h \\ \mathbf{e}^u \end{bmatrix}, \qquad \text{(Structure)} \tag{2.69a}$$

$$\begin{bmatrix} \partial_t h \\ \partial_t \mathbf{u} \end{bmatrix} = \begin{bmatrix} -f^h \\ -\mathbf{f}^u - \zeta \mathbf{u}^\perp \end{bmatrix}, \qquad \text{(Dynamics)} \tag{2.69b}$$

$$\begin{bmatrix} e^h \\ \mathbf{e}^u \end{bmatrix} = \begin{bmatrix} \delta_h H \\ \delta_u^T H \end{bmatrix}, \qquad \text{(Constit. Eq.)} \tag{2.69c}$$

where $e^h = \frac{1}{2}\mathbf{u} \cdot \mathbf{u} + g(h+b)$ expresses the hydrodynamic pressure function and $\mathbf{e}^u = h\mathbf{u}$ is the specific discharge vector field. Note that while the structure equation (2.69a) and the definition of co-state variables (2.69c) coincide with (2.48a) and (2.48c), the difference to the linear wave equation lies in (i) the non-quadratic energy functional (2.68) and (ii) the additional term $\zeta \mathbf{u}^\perp$ in the differential equation (2.69b) for $\mathbf{u}$.

Remark 2.8. As an alternative, the vorticity term can be added to the structure equations, which leads to a *non-canonical* differential operator, see e. g. [151] or [33], Section 6.2.

[20] It stems from the rotational part of the transport term in the momentum equation.

2.3.2.2 Exterior Calculus Notation

In order to write the equations in terms of differential forms, we first define

$$p := *h \in \Lambda^2(\Omega), \qquad q := \mathbf{u}^\flat = u\,dx + v\,dy \in \Lambda^1(\Omega) \tag{2.70}$$

as state differential forms, accordingly $f^p := *f^h \in \Lambda^2(\Omega)$ and $f^q := (\mathbf{f}^u)^\flat \in \Lambda^1(\Omega)$. With[21]

$$(\mathbf{u}^\perp)^\flat = -v\,dx + u\,dy = *q \in \Lambda^1(\Omega), \tag{2.71}$$

the effort differential forms

$$e^p = e^h \in \Lambda^0(\Omega), \qquad e^q = (-1)^{2-1} *(\mathbf{e}^u)^\flat \in \Lambda^1(\Omega), \tag{2.72}$$

exploiting the relations $\operatorname{grad} e^h = (\mathrm{d}e^h)^\sharp$, $\operatorname{div} \mathbf{e}^u = *\mathrm{d}(*(\mathbf{e}^u)^\flat)$, and with $**\lambda = (-1)^{k(n-k)}\lambda$ for some $\lambda \in \Lambda^k(\Omega)$, (2.69a)–(2.69c) can be expressed as

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & -\mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix}, \qquad \text{(Structure)} \tag{2.73a}$$

$$\begin{bmatrix} \partial_t p \\ \partial_t q \end{bmatrix} = \begin{bmatrix} -f^p \\ -f^q - \zeta * q \end{bmatrix}, \qquad \text{(Dynamics)} \tag{2.73b}$$

$$\begin{bmatrix} e^p \\ e^q \end{bmatrix} = \begin{bmatrix} \delta_p H \\ \delta_q H \end{bmatrix}. \qquad \text{(Constit. Eq.)} \tag{2.73c}$$

It is a straightforward exercise to verify that the effort differential forms defined in (2.72) (including the minus sign in e^q), i. e.

$$e^p = \frac{1}{2} p * p + g(*p + b), \qquad e^q = -*(h\mathbf{u})^\flat = hv\,dx - hu\,dy = -*p\,*q, \tag{2.74}$$

are indeed variational derivatives according to (2.40) of the non-quadratic functional

$$H = \int_\Omega \mathcal{H}, \qquad \mathcal{H} = \frac{1}{2} * p\, q \wedge *q + \frac{1}{2} g p * p + g p b. \tag{2.75}$$

The energy balance, see (2.61), is

$$\begin{aligned} \dot{H} &= \langle \delta_p H | \partial_t p \rangle_\Omega + \langle \delta_q H | \partial_t q \rangle_\Omega \\ &= \langle e^p | - f^p \rangle_\Omega + \langle e^q | - f^q - \zeta * q \rangle_\Omega. \end{aligned} \tag{2.76}$$

The only difference to the linear wave equation is the term

$$\langle e^q | - \zeta * q \rangle_\Omega = \int_\Omega \zeta * p\, (*q \wedge *q) = 0, \tag{2.77}$$

which does not contribute[22] to the energy balance. The result after integration by parts is again, cf. (2.62),

$$\partial_t H = (-1)^2 \langle \operatorname{tr} e^q | \operatorname{tr} e^p \rangle_{\partial\Omega}. \tag{2.78}$$

[21] The Hodge star applied to the basis 1-forms in 2D gives $*dx = dy$, $*dy = -dx$.

[22] $\lambda \wedge \lambda = 0$ for every differential form λ of *odd* degree.

2.3.3 Heat Equation

While the first examples are *hyperbolic* systems, i. e. equations that describe the linear/nonlinear propagation of waves, the *heat equation*

$$\partial_t x(\mathbf{z},t) = k\Delta x(\mathbf{z},t), \quad k>0, \quad \mathbf{z}\in\Omega \tag{2.79}$$

belongs to the class of *parabolic* systems[23].

2.3.3.1 Vector Calculus Notation

We start with its derivation from the conservation of internal energy on a constant volume $\Omega \subset \mathbb{R}^3$ of an incompressible, homogeneous medium. The first law of thermodynamics states that

$$\dot{U}(t) = \dot{Q}(t), \tag{2.80}$$

where U denotes the total internal energy on Ω and $\dot{Q}$ the heat flow into Ω. With u the internal energy density and $\mathbf{J}_Q \in \mathbb{R}^3$ the *heat flux* vector field, the balance equation becomes

$$\int_\Omega \partial_t u(\mathbf{z},t)\, d\mathbf{z} = \int_{\partial\Omega} -\mathbf{J}_Q(\mathbf{z},t)\cdot\mathbf{n}\, dS, \tag{2.81}$$

where $\mathbf{n}$ is the outer normal unit vector on $\partial\Omega$ and dS the infinitesimal surface element. Assuming constant density ρ (constant specific volume $v = 1/\rho$), the left hand term can be written

$$\int_\Omega \partial_t u(\mathbf{z},t)\, d\mathbf{z} = \int_\Omega \partial_T u|_v\, \partial_t T(\mathbf{z},t)\, d\mathbf{z}. \tag{2.82}$$

$\partial_T u|_v =: c_v(T)$ is the *specific heat capacity at constant volume*. Applying Gauss' divergence theorem, the right hand side of (2.81) becomes

$$\int_{\partial\Omega} -\mathbf{J}_Q(\mathbf{z},t)\cdot\mathbf{n}\, dS = -\int_\Omega \operatorname{div}\mathbf{J}_Q(\mathbf{z},t)\, d\mathbf{z}. \tag{2.83}$$

Identifying the right hand side integrands of the last two equations, we obtain

$$c_v(T)\partial_t T(\mathbf{z},t) = -\operatorname{div}\mathbf{J}_Q(\mathbf{z},t). \tag{2.84}$$

With the temperature gradient as the *thermodynamic driving force*, the heat flux can be described by *Fourier's law*

$$\mathbf{J}_Q(\mathbf{z},t) = -\lambda(T)\ \operatorname{grad} T(\mathbf{z},t), \tag{2.85}$$

[23] For the classification of PDEs into hyperbolic, parabolic and elliptic equations, see e. g. [144], Section 4.4.

where $\lambda(T)$ denotes the *heat conductivity*. Putting the pieces together, we obtain

$$\partial_t T(\mathbf{z},t) = \frac{1}{c_V(T)} \operatorname{div}(\lambda(T)\ \operatorname{grad} T(\mathbf{z},t)), \tag{2.86}$$

which for the case of constant heat capacity and conductivity can be written in the form (2.79).

The conservation of internal energy, the definition of the driving force $\mathbf{F}$, as well as the calorimetric equation and Fourier's law can be represented by

$$\begin{bmatrix} f^u \\ \mathbf{F} \end{bmatrix} = \begin{bmatrix} 0 & \operatorname{div} \\ \operatorname{grad} & \mathbf{0} \end{bmatrix} \begin{bmatrix} T \\ \mathbf{J}_Q \end{bmatrix}, \qquad \text{(Structure)} \tag{2.87a}$$

$$\partial_t u = -f^u, \qquad \text{(Dynamics)} \tag{2.87b}$$

$$\begin{bmatrix} T \\ \mathbf{J}_Q \end{bmatrix} = \begin{bmatrix} \frac{1}{c_v(T)} u \\ -\lambda(T)\mathbf{F} \end{bmatrix}. \qquad \text{(Constit. Eq.)} \tag{2.87c}$$

2.3.3.2 Exterior Calculus Notation

Now assume an n-dimensional spatial domain Ω, $n \in \{1,2,3\}$. By defining the differential forms of the internal energy density $p = *u \in \Lambda^n(\Omega)$, the temperature $e^p = T \in \Lambda^0(\Omega)$ (function), the driving force $f^q = (\operatorname{grad} T)^\flat \in \Lambda^1(\Omega)$ and the heat flux[24] $e^q = (-1)^{n-1} * \mathbf{J}_Q^\flat \in \Lambda^{n-1}(\Omega)$, Eqs. (2.87a)–(2.87c) can be rewritten in the form

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & (-1)^{n-1}\mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix}, \qquad \text{(Structure)} \tag{2.88a}$$

$$\partial_t p = -f^p, \qquad \text{(Dynamics)} \tag{2.88b}$$

$$\begin{bmatrix} e^p \\ e^q \end{bmatrix} = \begin{bmatrix} \frac{1}{c_v(e^p)} * p \\ (-1)^n \lambda(e^p) * f^q \end{bmatrix}. \qquad \text{(Constit. Eq.)} \tag{2.88c}$$

As in the case of the wave equation, the formal skew-adjointness of the matrix operator in (2.88a) imposes the *structural balance equation* (2.60), i. e.

$$\langle e^p | f^p \rangle_\Omega + \langle e^q | f^q \rangle_\Omega + (-1)^n \langle \operatorname{tr} e^q | \operatorname{tr} e^p \rangle_{\partial\Omega} = 0. \tag{2.89}$$

For the given choice of $e^p = T$, this equation is not an energy nor an entropy balance. It becomes an entropy balance if the reciprocal temperature is chosen as intensive variable, $e^p = \frac{1}{T}$. In this case, e^p can be expressed as variational derivative of the *entropy functional* S with respect to the internal energy density: $\frac{1}{T} = \delta_u S$, see [50], Section 4.2.2. While with $f^p = -\partial_t u$, the expression $\langle e^p | f^p \rangle_\Omega = -\partial_t S$ represents the change of entropy, $\langle e^q | f^q \rangle_\Omega$ is non-negative and stands for entropy generation.

Comparing (2.79) with (2.46), the only difference is the order of the time derivative, which causes the completely different nature of the solutions of the

[24] Equivalently, $*e^q = (\mathbf{J}_Q)^\flat \in \Lambda^1(\Omega)$.

hyperbolic wave and the *parabolic* heat equation. While the former is a superposition of travelling waves (transporting initial and boundary conditions), initial and boundary data is instantaneously smoothened in the solution of the heat equation. In Chapters 4 and 6, we will take into account this fundamental difference in the parametrization of the proposed structure-preserving discretization scheme. Moreover, for the wave equation, simulation and control design in *discrete time* using an appropriate geometric numerical integration scheme, see Chapter 5, allows to properly reproduce the unsmoothened propagation of initial and boundary data.

Chapter 3

Discrete Port-Hamiltonian Formulation of Conservation Laws

The chapter[1] deals with the *direct discrete formulation* of systems of two conservation laws on *dual chain complexes* in port-Hamiltonian form. Based on integral balance equations and topological information, this representation of the infinite-dimensional system is *exact* and qualifies as a control model. For simulation, feedforward control and observer design, a *consistent* numerical approximation is required, which yields a *discretized energy*, from which the lumped constitutive equations are derived. We refer to the previous works [175], [174] and extend their results to cases of an *arbitrary*, non-uniform distribution of different types of boundary inputs (in the sense of the *causality* of the boundary ports).

After the integral representation of a canonical hyperbolic system of two conservation laws in Section 3.1, we provide an overview of some concepts from *discrete exterior calculus* in Section 3.2. They are necessary to describe the *oriented topological objects* and their relations, on which the conservation laws are evaluated. Section 3.3 describes the systematic construction of the primal and the dual grid/complex in order to incorporate different types of boundary inputs (corresponding to Dirichlet or Neumann conditions). The result is an explicit finite-dimensional system representation in terms of integral states and integral efforts, which features both types of boundary inputs. The structure of the discrete models reveals the slight inexactitude of power-conjugation and collocation of the lumped boundary power variables, which is a typical feature of approximation methods on dual meshes. In Section 3.4, we consider the numerical approximation of the constitutive equations, which is the bridge to classical finite volume (control volume) schemes. By means of the 2D example of the nonlinear, irrotational shallow water equations on rectangular grids, we show that the numerical approximation of the flux functions (integral efforts) by a centered scheme is consistent of order 2 in the interior, which extends the

[1] Up to some minor corrections, this chapter corresponds to the main part of the journal article [103].

1D result in [95]. The shifted grids require to assume values for different state variables along the boundary. We discuss the errors due to an (in-)consistent assignment of these *ghost values* and give a series of remarks. We conclude the chapter with references to related recent works and some notes on the current application of the presented approach to complex heterogeneous systems in Section 3.5.

3.1 Discrete Representation of Conservation Laws

We consider (lossless) systems of two conservation laws on an n-dimensional open domain Ω with Lipschitz boundary $\partial\Omega$. An integral representation, see [50], Section 4.2.1, is given by

$$\frac{d}{dt}\int_{c_p} p + \int_{\partial c_p} \beta^p = 0, \qquad \frac{d}{dt}\int_{\hat{c}_q} q + \int_{\partial \hat{c}_q} \beta^q = 0, \tag{3.1}$$

where c_p and $\hat{c}_q$ are p- and q-dimensional subsets of Ω. $p \in \Lambda^p(\Omega)$, $q \in \Lambda^q(\Omega)$ are differential forms[2] that represent the conserved quantities, and $\beta^p \in \Lambda^{p-1}(\Omega)$, $\beta^p \in \Lambda^{q-1}(\Omega)$ denote *fluxes*. In a canonical PH system of two conservation laws, as introduced in Definition 2.4, the relation $p + q = n + 1$ between the degrees of the differential forms and the dimension of Ω holds. Moreover, the fluxes are determined from *effort* or co-state differential forms according to

$$\begin{bmatrix} \beta^p \\ \beta^q \end{bmatrix} = \begin{bmatrix} 0 & (-1)^{pq+1} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix}, \tag{3.2}$$

where

$$e^p = \delta_p H \in \Lambda^{n-p}(\Omega) \qquad \text{and} \qquad e^q = \delta_q H \in \Lambda^{n-q}(\Omega) \tag{3.3}$$

are derived from a Hamiltonian functional $H(p,q) = \int_\Omega \mathcal{H}(p,q,z)$ with the Hamiltonian density n-form $\mathcal{H}$, see (2.38), (2.39). Merging (3.1) with (3.2) yields

$$\begin{aligned} \frac{d}{dt}\int_{c_{p,i}} p &= (-1)^{pq}\int_{\partial c_{p,i}} e^q, \\ \frac{d}{dt}\int_{\hat{c}_{q,j}} q &= -\int_{\partial \hat{c}_{q,j}} e^p, \end{aligned} \tag{3.4}$$

which is the basis for the discrete modeling approach presented in this chapter. The indices indicate that we will evaluate both conservation laws on *discrete geometric objects* $c_{p,i}$ and $\hat{c}_{q,j}$ of the corresponding discretization mesh. We will

[2] Unlike the mixed Galerkin approach in Chapter 4, we do not approximate the spatial distribution of the differential forms. We therefore do not need to choose the compatible functional spaces for the approximation. Instead, we assume all finite integrals to be bounded, and all restrictions of differential forms to boundaries to be well-defined, which is the case for smooth differential forms.

use *topologically dual* or *staggered* meshes (a "hat" denotes an object on the dual mesh) in order to account for the different physical nature of the conserved quantities p and q, and also to be able to discretize the constitutive equations in a direct, consistent manner. The sequences of discrete topological objects (oriented volumes, faces, edges and nodes), together with the boundary maps, define so-called *chain complexes* on both meshes. Therefore, we refer to both meshes over an n-dimensional domain Ω with Lipschitz boundary $\partial\Omega$ as *primal* and *dual* n*-complex*.

Remark 3.1. In this chapter, we directly consider the conservation laws (including the dynamics) on dual n-complexes. For the notion of canonical Dirac structures on n-complexes, we refer to [198]. We frequently refer to [175], [174], where modeling of conservation laws using the "discrete exterior geometry approach" has been presented for control inputs of *uniform* physical type, and without treating the numerical approximation, in particular for nonlinear systems.

3.2 Preliminaries from Discrete Exterior Calculus

We summarize and explain the basic ideas from *discrete exterior calculus*, see e. g. [84], [47], and on topological duality, as they are used in the direct discrete modeling of canonical PH systems of conservation laws in [175], [174].

3.2.1 Oriented Discretization Mesh

To introduce the necessary notions for the integration of differential forms on *discrete* objects, we consider Fig. 3.1 which we will identify as the graphical representation of a 2-complex. The non-rectangular, *oriented* mesh could result from a polyhedral tessellation[3] K on a subset of $\mathbb{R}^2$. The mesh is based on five nodes $n_1, \ldots, n_5 \in \mathcal{N}$, connected by six oriented edges $e_1, \ldots, e_6 \in \mathcal{E}$ which divide the convex hull of the nodes into two oriented faces $f_1, f_2 \in \mathcal{F}$. The sets of nodes, edges and faces have cardinalities $|\mathcal{N}| = 5$, $|\mathcal{E}| = 6$ and $|\mathcal{F}| = 2$.

Remark 3.2. We have arbitrarily defined the object in Fig. 3.1 to represent a mesh on $\mathbb{R}^2$. It could also display some "folded" two-dimensional manifold in $\mathbb{R}^3$. To know about the shape of the underlying object, the *topological* information must be completed by *geometric* data.

3.2.2 Cells, Chains and Cochains

We start with the definition of the most important discrete objects of a *chain complex*.

[3] To distinguish from a *simplicial triangulation* as in [174].

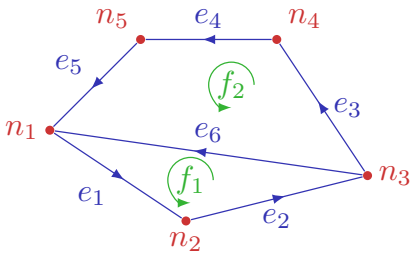

Figure 3.1: A non-simplicial mesh in 2D, composed of oriented cells.

Definition 3.1 ([8], p. 184)**.** A *j-dimensional cell* or *j-cell* of an n-dimensional smooth manifold M is characterized by an oriented convex polyhedron $D \subset \mathbb{R}^j$, and a differentiable map[4] $f : D \to M$.

Nodes, edges and faces in Fig. 3.1 represent 0-cells, 1-cells and 2-cells, with orientations indicated by the arrows. The sets $\mathcal{N}$, $\mathcal{E}$ and $\mathcal{F}$ (or subsets thereof) are the bases of *j-chains* ($j = 0, 1, 2$) according to the following definition.

Definition 3.2 ([8], p. 185)**.** A *j-dimensional chain* or *j-chain* is a finite-dimensional collection (or a formal sum) of j-cells σ_i, weighted by scalars (multiplicities) m_i: $c_j = m_1\sigma_1 + \ldots + m_r\sigma_r$. The linear vector space of j-chains on a tessellation K is denoted[5] $C_j(K;\mathbb{R})$.

According to the definition, the simplest j-chain is a j-cell. If the multiplicities are restricted to $\{-1, 0, 1\}$, a j-chain can be considered as the concatenation of several j-cells, e. g. the 1-chain $e_3 + e_4 + e_5 - e_6 =: \partial_2 f_2$, which forms the (oriented) boundary of the 2-cell f_2.

Definition 3.3 ([65], p. 638)**.** A *j-cochain* is a linear functional on the j-chains.

The linear functional on the j-chains can be understood via the *duality pairing* $\langle \cdot, \cdot \rangle : C^j(K;\mathbb{R}) \times C_j(K;\mathbb{R}) \to \mathbb{R}$, where $C^j(K;\mathbb{R})$ denotes the linear vector space of cochains. Hence, j-cochains $c^j \in C^j(K;\mathbb{R})$ are *algebraically*[6] dual objects with respect to this pairing. In our context, j-cochains will contain the integral values of j-forms on j-cells. Later on, we will define discrete state and effort vectors which can be understood as vector-valued representations of such j-cochains.

[4] This map is the identity function $f = \mathrm{id}$ if $M = \mathbb{R}^n$.
[5] $\mathbb{R}$ indicates that the multiplicities are real-valued.
[6] In contrast to topological duality based on which the dual grid/complex is constructed.

3.2.3 Boundary Maps and Primal Chain Complex

Each j-cell has a *boundary* which is composed of $(j-1)$-cells whose orientation is *induced* by the orientation of the j-cell, see e. g. $\partial_2 f_2$ as defined above. The symbol ∂_j will at the same time denote the boundary map $\partial_j : C_j \to C_{j-1}$ and its matrix representation[7]. Let $\mathbf{u}_k^{\mathcal{F}}$ ($\mathbf{u}_l^{\mathcal{E}}$) be an $|\mathcal{F}|$-dimensional ($|\mathcal{E}|$-dimensional) unit vector with 1 at the k-th (l-th) position. Then $\boldsymbol{\partial}_2\mathbf{u}_k^{\mathcal{F}}$ ($\boldsymbol{\partial}_1\mathbf{u}_l^{\mathcal{E}}$) returns the $|\mathcal{E}|$-dimensional ($|\mathcal{N}|$-dimensional) vector with elements from the set $\{-1, 0, 1\}$ indicating the edges (nodes) that form the boundary of the 2-cell f_k (the 1-cell e_l). The *boundary* or *incidence*[8] *matrices* for the example in Fig. 3.1 are

$$\boldsymbol{\partial}_1 = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}, \qquad \boldsymbol{\partial}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}. \tag{3.5}$$

It is easy to verify at the example that $\partial_1 \circ \partial_2 = 0$, i. e. the range of ∂_2 spans the kernel of ∂_1. This property holds for any concatenation of two subsequent boundary maps: "*[T]he boundary of each chain itself has zero boundary*" ([59], p. 59) and is known in general as the *complex property*. We can illustrate the relations between the spaces of j-chains and the boundary maps with $\partial_{j-1} \circ \partial_j = 0$, $j = 2, \ldots, n$, by the sequence diagram

$$C_n(K;\mathbb{R}) \xrightarrow{\partial_n} C_{n-1}(K;\mathbb{R}) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} C_0(K;\mathbb{R}). \tag{3.6}$$

Figure 3.1 represents such a *chain complex* for $n = 2$. By identifying the graphical representation with the object behind, we will for brevity refer to it as a 2-complex.

Remark 3.3. A *(chain) complex* is in a general manner defined as a sequence of abelian groups (e. g. vector spaces), connected by homomorphisms, i. e. mappings that preserve the group operation[9]. Famous examples are (i) the sequence of smooth scalar- and vector-valued function spaces, connected via the differential operations grad/rot/div and (ii) the so-called *de Rham* complex with spaces of (smooth) differential forms, connected via the exterior derivative d.

3.2.4 Coboundary Maps and Cochain Complex

Using the duality pairing between chains and cochains, the *coboundary operator* d^j can be defined via

$$\langle c^{j-1}, \partial_j c_j \rangle = \langle \mathrm{d}^j c^{j-1}, c_j \rangle, \tag{3.7}$$

[7] It will be typeset in boldface, if the matrix representation is emphasized.

[8] As in [198], we will use rather the terms (co-)incidence maps instead of (co-)boundary maps to distinguish from the boundary port variables.

[9] See e. g. [90], p. 127.

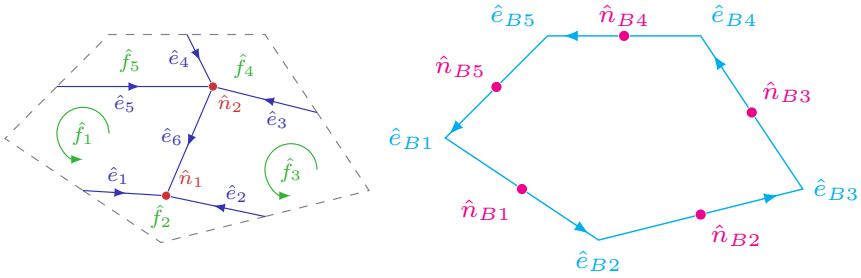

Figure 3.2: The dual 2-complex and the dual boundary.

which gives rise to the sequence diagram

$$C^0(K;\mathbb{R}) \xrightarrow{\mathrm{d}^1} C^1(K;\mathbb{R}) \xrightarrow{\mathrm{d}^2} \quad \cdots \quad \xrightarrow{\mathrm{d}^n} C^n(K;\mathbb{R}) \tag{3.8}$$

of the *cochain complex* with $\mathrm{d}^j \circ \mathrm{d}^{j-1} = 0$, $j = 2, \ldots, n$. Assuming the chain c_j represented by a column vector and the cochain c^{j-1} by a row vector, the relation between the matrix representations of boundary and coboundary map (or incidence and co-incidence matrix, respectively) becomes evident from (3.7):

$$\mathbf{d}^j = (\boldsymbol{\partial}_j)^T. \tag{3.9}$$

Remark 3.4. The co-incidence operator d^j is the discrete counterpart of the exterior derivative, and therefore can be understood as *discrete exterior derivative* [174]. Accordingly, j-cochains in the discrete setting correspond naturally to j-forms and Eq. (3.7) can be considered the discrete version of Stokes' theorem[10].

3.2.5 Trace Operators

The *trace operators* $\mathrm{tr}_j : C_j(K;\mathbb{R}) \to C_j(\partial K, \mathbb{R})$, $j = 0, \ldots, n-1$, isolate the j-chains on the boundary of the n-complex. For the example of the 2-complex depicted in Fig. 3.1, and again identifying the operator with its matrix representation, we have

$$\mathbf{tr}_0 = \mathbf{I}_5, \qquad \mathbf{tr}_1 = \begin{bmatrix} \mathbf{I}_5 & \mathbf{0}_{5\times 1} \end{bmatrix}. \tag{3.10}$$

3.2.6 The Dual n-Complex

To each j-cell on an n-complex, we can associate a *topologically* dual $(n-j)$-cell, which can have different *geometric* realizations (e. g. barycentric, circumcentric). In our 2D example, a node has a dual surrounding face, an edge has a

[10] Consider the generalized Stokes' theorem A.1, $\int_{\partial\Omega} \omega = \int_{\Omega} \mathrm{d}\omega$, and rewrite it as a pairing of the differential form with the integration domain: $\langle \omega, \partial\Omega \rangle = \langle \mathrm{d}\omega, \Omega \rangle$.

dual edge across it, see Fig. 3.2, left. When we refer to "topological duality", which manifests itself in the relation of primal and dual (co-)incidence matrices, see below, we tacitly associate it with its geometric realization.

3.2.7 Duality Relations of the Co-Incidence Matrices

Between the co-incidence matrices of an n-complex and its dual, the following relation holds[11] (first formula in Section 3.3 of [174], in our notation):

$$\hat{\mathbf{d}}^{n-j+1} = (-1)^j (\mathbf{d}^j)^T. \tag{3.11}$$

In our example, we have $\hat{\mathbf{d}}^1 = (\mathbf{d}^2)^T$ and $\hat{\mathbf{d}}^2 = -(\mathbf{d}^1)^T$. The construction of dual $(n-j)$-cells leaves out the boundary ∂K of the original complex, which is represented by a boundary $(n-1)$-chain. The dual n-complex is provided with a boundary $(n-1)$-chain (whose j-cells are indexed B) by topological duality on ∂K. With the second formula in Section 3.3 of [174] in our notation,

$$\hat{\mathbf{d}}_B^{n-j} = (-1)^j (\mathbf{tr}_j)^T, \tag{3.12}$$

we obtain in our example $\hat{\mathbf{d}}_B^1 = -(\mathbf{tr}_1)^T$ and $\hat{\mathbf{d}}_B^2 = (\mathbf{tr}_0)^T$.

3.3 Discrete Conservation Laws on n-Complexes

We study systems of two conservation laws in integral PH form (3.4) on an n-dimensional domain $\Omega \subset \mathbb{R}^n$. Before we state our general result for

$$p = n \in \{1, 2, 3\}, \quad q = 1, \tag{3.13}$$

we illustrate the approach, in particular the construction of the dual complexes, for $n = 2$ and rectangular grids.

3.3.1 Non-Uniform Boundary Inputs

Constructing the dual n-complexes as sketched in the previous section, we can observe the following concerning the inputs for the integral PH formulation of the conservation laws. If *all the* n-cells on the primal complex are integration domains for the conserved quantity $p \in \Lambda^n(\Omega)$, then the boundary inputs will be *exclusively* related to the $(n-1)$-cochain representing the integrals of $\operatorname{tr} e^q \in \Lambda^{n-1}(\partial\Omega)$ on the boundary. Correspondingly for $q \in \Lambda^1(\Omega)$ on the dual 1-cells and the 0-cochain of boundary values of $\operatorname{tr} e^p \in \Lambda^0(\partial\Omega)$. This situation of a *uniform causality* of the boundary ports is treated in the previous works [198] and [174].

However, in most practical cases for modeling and control, a given physical variable will play the role of an input on parts of the boundary, while its power-conjugate will be the input on the rest. The causality along the boundary

[11] We denote all quantities on the dual complex (in particular the dual n-cells and the (co-)incidence maps) with a "hat".

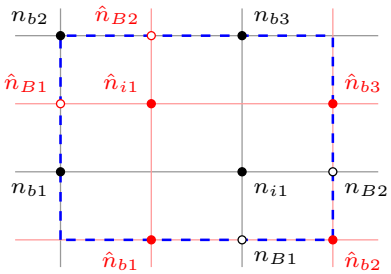

Figure 3.3: Definition of nodes on the primal and dual 2-complex.

depends on the boundary conditions which shall be imposed. This situation is designated *non-uniform (distribution of) inputs* and must be accounted for in the formulation of the dual complexes relative to the system boundary.

3.3.2 Construction of the Dual Complexes

Based on two given staggered meshes and the system boundary, we will construct two n-complexes, the first one representing the integration domains (n-cells) for $p \in \Lambda^n(\Omega)$ and their boundary[12] $(n-1)$-cells, associated to $e^q \in \Lambda^{n-1}(\Omega)$. The second (or dual) n-complex contains the integration domains (1-cells) for $q \in \Lambda^1(\Omega)$ and their boundaries (0-cells) at which the function values of $e^p \in \Lambda^0(\Omega)$ are evaluated. To define the different subsets of j-cells, related to state and co-state variables, boundary in- and outputs along $\partial\Omega$, we exploit topological duality on both complexes and the boundary. The classification of cells on the primal and dual complex, as illustrated below for $n = 2$, can be adapted in a straightforward manner for the other cases $n = 1$ and $n = 3$.

3.3.2.1 Example: Two-Dimensional Rectangular Grids

We adopt the notation from the previous section with a "hat" for quantities on the dual complex and refer to 0-/1-/2-cells as nodes/edges/faces. Figure 3.3 shows two staggered rectangular grids (primal: black, dual: red) with their nodes. The system boundary (dashed blue) coincides everywhere with either a line of the primal or the dual mesh. The classification of primal and dual j-cells, as introduced below, is illustrated in Figs. 3.4 and 3.5.

Classification of j-cells. We classify the different sets of j-cells as follows. The indices i and b denote *interior* and *boundary* objects that are constructed based on nodes of the primal and the dual grid that lie *within* or *on* the system

[12] Here, we mean the boundary of the n-cells, not the boundary of Ω.

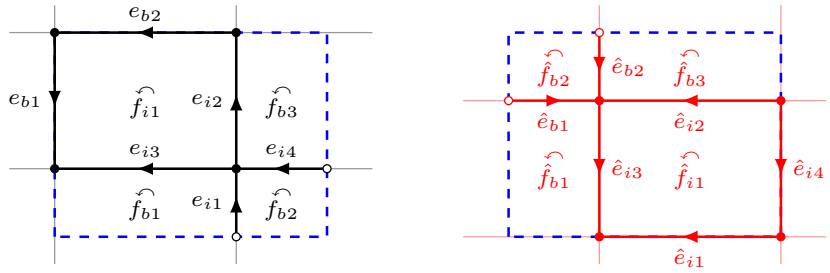

Figure 3.4: Interior and boundary edges.

boundary. The capital letter B refers to *additional* (complementary) boundary nodes and edges.

1. **Nodes.**

   $n_{i,\cdot} \in \mathcal{N}_i$ and $n_{b,\cdot} \in \mathcal{N}_b$: Nodes of the primal mesh, within and on the system boundary.

   $\hat{n}_{i,\cdot} \in \hat{\mathcal{N}}_i$ and $\hat{n}_{b,\cdot} \in \hat{\mathcal{N}}_b$: Nodes of the dual mesh, within and on the system boundary.

2. **Additional boundary nodes.**

   $n_{B,\cdot} \in \mathcal{N}_B$ and $\hat{n}_{B,\cdot} \in \hat{\mathcal{N}}_B$: Additional nodes on the intersection of interior edges and the system boundary.

3. **Primal edges.**

   $e_{i,\cdot} \in \mathcal{E}_i$ and $e_{b,\cdot} \in \mathcal{E}_b$, *interior* and *boundary* edges: Connect the above-defined primal nodes and lie *within* and *on* the system boundary, respectively.

4. **Dual edges.**

   $\hat{e}_{i,\cdot} \in \hat{\mathcal{E}}_i$ and $\hat{e}_{b,\cdot} \in \hat{\mathcal{E}}_b$: Connect the above-defined dual nodes. Their indices follow from topological duality to the primal edges $e_{i,\cdot} \in \mathcal{E}_i$ and $e_{b,\cdot} \in \mathcal{E}_b$.

5. **Faces.**

   $\hat{f}_{i,\cdot} \in \hat{\mathcal{F}}_i$ and $\hat{f}_{b,\cdot} \in \hat{\mathcal{F}}_b$: Dual faces, topologically dual to primal nodes $n_{i,\cdot} \in \mathcal{N}_i$ and $n_{b,\cdot} \in \mathcal{N}_b$.

   $f_{i,\cdot} \in \mathcal{F}_i$ and $f_{b,\cdot} \in \mathcal{F}_b$: Primal faces, topologically dual to dual nodes $\hat{n}_{i,\cdot} \in \hat{\mathcal{N}}_i$ and $\hat{n}_{b,\cdot} \in \hat{\mathcal{N}}_b$.

6. **Additional boundary edges.**

   $e_{B,\cdot} \in \mathcal{E}_B$ and $\hat{e}_{B,\cdot} \in \hat{\mathcal{E}}_B$: Edges (more precisely 1-chains, as they run over the corners of the rectangle) on the system boundary that complete the boundaries of the faces $f_{b,\cdot} \in \mathcal{F}_b$ and $\hat{f}_{b,\cdot} \in \hat{\mathcal{F}}_b$.

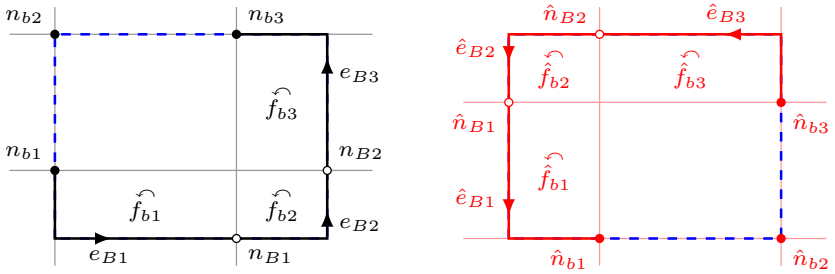

Figure 3.5: Additional or complementary boundary edges.

Table 3.1 shows the cardinalities of the defined sets of j-cells and j-chains on both complexes, and thereby illustrates their duality. By the proposed construction, the following objects are topologically dual *on the boundary*: $\hat{n}_{b,\cdot}$ and $e_{B,\cdot}$; $\hat{n}_{B,\cdot}$ and $e_{b,\cdot}$; $n_{b,\cdot}$ and $\hat{e}_{B,\cdot}$; $n_{B,\cdot}$ and the edges $\hat{e}_{i,\cdot}$ *on* the system boundary.

Incidence matrices. The primal and dual 2-complex in the example have the following incidence matrices (faces to edges and edges to nodes), for which the complex property $\partial_1 \circ \partial_2 = 0$ and $\hat{\partial}_1 \circ \hat{\partial}_2 = 0$ can be immediately verified:

$$\boldsymbol{\partial}_2 = \left[\begin{array}{c:ccc} 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ \hdashline 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right], \tag{3.14a}$$

$$\boldsymbol{\partial}_1 = \left[\begin{array}{cccc:cc:ccc} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ \hline -1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & -1 \end{array}\right], \tag{3.14b}$$

Table 3.1: Cardinalities of the sets of primal and dual topological objects in the example.

Primal	$\mathcal{N}_i$	$\mathcal{N}_b$	$\mathcal{N}_B$	$\mathcal{F}_i$	$\mathcal{F}_b$, $\mathcal{E}_B$		$\mathcal{E}_i$	$\mathcal{E}_b$
Dual	$\hat{\mathcal{F}}_i$	$\hat{\mathcal{F}}_b$, $\hat{\mathcal{E}}_B$		$\hat{\mathcal{N}}_i$	$\hat{\mathcal{N}}_b$	$\hat{\mathcal{N}}_B$	$\hat{\mathcal{E}}_i$	$\hat{\mathcal{E}}_b$
#	1	3	2	1	3	2	4	2

$$\hat{\boldsymbol{\partial}}_2 = \left[\begin{array}{c:ccc} -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ \hdashline 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right], \tag{3.15a}$$

$$\hat{\boldsymbol{\partial}}_1 = \left[\begin{array}{ccc:cc:ccc} 0 & 1 & -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ \hdashline 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ \hline 0 & 0 & 0 & 0 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 1 \end{array}\right]. \tag{3.15b}$$

Incidence submatrices. The incidence matrices $\boldsymbol{\partial}_j$, $j = 1, 2$, on the primal complex are partitioned according to the categories of involved j- and $(j-1)$-chains,

$$\boldsymbol{\partial}_2 = \left[\begin{array}{cc} \boldsymbol{\partial}_2^{ii} & \boldsymbol{\partial}_2^{ib} \\ \hline \boldsymbol{\partial}_2^{bi} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right], \qquad \boldsymbol{\partial}_1 = \left[\begin{array}{cc:c} \boldsymbol{\partial}_1^{ii} & \mathbf{0} & \mathbf{0} \\ \hline \boldsymbol{\partial}_1^{bi} & \boldsymbol{\partial}_1^{bb} & \boldsymbol{\partial}_1^{bB} \\ \hline \boldsymbol{\partial}_1^{Bi} & \mathbf{0} & \boldsymbol{\partial}_1^{BB} \end{array}\right]. \tag{3.16}$$

The columns of $\boldsymbol{\partial}_2^{bi}$, for instance, represent the boundary edges of the inner faces that lie on the system boundary. The columns of $\boldsymbol{\partial}_1^{bB}$ and $\boldsymbol{\partial}_1^{BB}$ represent the terminal nodes of additional boundary edges, which are boundary nodes in the former and additional boundary nodes in the latter case. The zero matrices and the identity matrix result from the construction of the subsets[13] indexed i, b and B.

On the dual complex, the structure of the incidence matrices is

$$\hat{\boldsymbol{\partial}}_2 = \left[\begin{array}{cc} \hat{\boldsymbol{\partial}}_2^{ii} & \hat{\boldsymbol{\partial}}_2^{ib} \\ \hline \mathbf{0} & \hat{\boldsymbol{\partial}}_2^{bb} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right], \quad \hat{\boldsymbol{\partial}}_1 = \left[\begin{array}{cc:c} \hat{\boldsymbol{\partial}}_1^{ii} & \hat{\boldsymbol{\partial}}_1^{ib} & \mathbf{0} \\ \hline \hat{\boldsymbol{\partial}}_1^{bi} & \mathbf{0} & \hat{\boldsymbol{\partial}}_1^{bB} \\ \hline \mathbf{0} & -\mathbf{I} & \hat{\boldsymbol{\partial}}_1^{BB} \end{array}\right]. \tag{3.17}$$

The different locations of the zero matrices result from the definition of dual objects. For example b-indexed edges on the dual complex do not lie on its

[13] $\boldsymbol{\partial}_2^{bb} = \mathbf{0}$ as the b-edges lie on the boundary of i-faces. $\boldsymbol{\partial}_2^{Bi} = \mathbf{0}$ as the B-edges lie on the boundary of the b-faces. $\boldsymbol{\partial}_2^{Bb} = \mathbf{I}$ by definition of the B-edges to complete the boundary of the positively oriented b-faces. $\boldsymbol{\partial}_1^{ib} = \mathbf{0}$ and $\boldsymbol{\partial}_1^{iB} = \mathbf{0}$ as i-nodes are at the terminals of i-edges only. $\boldsymbol{\partial}_1^{Bb} = \mathbf{0}$ as the b-edges are terminated only by b-nodes.

boundary. $\hat{\boldsymbol{\partial}}_1^{Bb} = -\mathbf{I}$ comes from the one-to-one relation of B-nodes and b-edges on the dual complex, and the orientation of the edges $\hat{e}_{b,\cdot}$, which is induced by the positive orientation of the primal boundary.

3.3.2.2 Duality Relations of Incidence Submatrices

Using the duality relations (3.11), (3.12), and based on an analogous construction of the dual complexes for $n \in \{1, 2, 3\}$, the following duality relations between the (co-)incidence submatrices can be given. Given the incidence matrices (on either the primal or the dual complex), the co-incidence matrices result from transposition:

$$\mathbf{d}_{\alpha\beta}^{j} = (\boldsymbol{\partial}_j^{\beta\alpha})^T, \qquad \alpha, \beta \in \{i, b, B\}. \tag{3.18}$$

The submatrices of $\boldsymbol{\partial}_j$ and $\mathbf{d}^j$ (upper or upper left 2×2 block matrices in the above example) that relate i and b indexed cells will be designated (i, b). As the relations between the i and b indexed cells are well-defined by topological duality, Eq. (3.11) applies accordingly:

$$\hat{\mathbf{d}}_{(i,b)}^{j} = (-1)^{n-j+1}(\mathbf{d}_{(b,i)}^{n-j+1})^T. \tag{3.19}$$

For the dual co-incidence matrices that relate b-indexed $(j-1)$-cells (in the interior) and B-indexed j-chains (on the boundary), the following holds:

$$\hat{\mathbf{d}}_{bB}^{j} = (-1)^{n-j}\mathbf{I}. \tag{3.20}$$

This relation corresponds to Eq. (3.12), where the trace matrix boils down to the identity matrix due to the fact that *all* B-indexed cells live on the boundary.

3.3.3 Discrete PH Representation

The integral conservation laws are now written in a compact form, exploiting the topological description of the primal and the dual mesh in terms of dual n-complexes. We introduce the following notation. $\mathbf{P}_i \in \mathbb{R}^{|\mathcal{F}_i|}$, $\mathbf{P}_b \in \mathbb{R}^{|\mathcal{F}_b|}$ and $\hat{\mathbf{Q}}_i \in \mathbb{R}^{|\hat{\mathcal{E}}_i|}$, $\hat{\mathbf{Q}}_b \in \mathbb{R}^{|\hat{\mathcal{E}}_b|}$ are vector representations of the primal 2-cochains and the dual 1-cochains that correspond to the integral conserved quantities on the primal 2-cells and the dual 1-cells. The elements of the vectors are

$$\begin{aligned} [\mathbf{P}_i]_j &= \int_{f_{i,j}} p, \qquad j = 1, \ldots, |\mathcal{F}_i|, \\ [\mathbf{P}_b]_k &= \int_{f_{b,k}} p, \qquad k = 1, \ldots, |\mathcal{F}_b|, \end{aligned} \tag{3.21}$$

and

$$\begin{aligned} [\hat{\mathbf{Q}}_i]_j &= \int_{\hat{e}_{i,j}} q, \qquad j = 1, \ldots, |\hat{\mathcal{E}}_i|, \\ [\hat{\mathbf{Q}}_b]_k &= \int_{\hat{e}_{b,k}} q, \qquad k = 1, \ldots, |\hat{\mathcal{E}}_b|. \end{aligned} \tag{3.22}$$

The vector representations $\mathbf{e}_i^q \in \mathbb{R}^{|\mathcal{E}_i|}$, $\mathbf{e}_b^q \in \mathbb{R}^{|\mathcal{E}_b|}$, $\mathbf{e}_B^q \in \mathbb{R}^{|\mathcal{E}_B|}$ and $\hat{\mathbf{e}}_i^p \in \mathbb{R}^{|\hat{\mathcal{N}}_i|}$, $\hat{\mathbf{e}}_b^p \in \mathbb{R}^{|\hat{\mathcal{N}}_b|}$, $\hat{\mathbf{e}}_B^p \in \mathbb{R}^{|\hat{\mathcal{N}}_B|}$ of the primal 1-cochains and the dual 0-cochains are defined accordingly. Their elements contain the integrals of the effort variables (which play the role of fluxes in the conservation laws) on the primal edges and their evaluations on the dual nodes.

For the "interior" integration domains we obtain, combining the integral representation (3.4) on all primal 2-chains and dual 1-chains,

$$\begin{aligned} \dot{\mathbf{P}}_i &= \mathbf{d}_{ii}^p (-1)^{pq} \mathbf{e}_i^q + \mathbf{d}_{ib}^p (-1)^{pq} \mathbf{e}_b^q \\ \dot{\hat{\mathbf{Q}}}_i &= -\hat{\mathbf{d}}_{ii}^q \hat{\mathbf{e}}_i^p - \hat{\mathbf{d}}_{ib}^q \hat{\mathbf{e}}_b^p. \end{aligned} \tag{3.23}$$

The relations between the discrete integration domains are expressed in terms of the primal and the dual co-incidence matrices. Accordingly for the "boundary" integration domains:

$$\begin{aligned} \dot{\mathbf{P}}_b &= \mathbf{d}_{bi}^p (-1)^{pq} \mathbf{e}_i^q + \mathbf{d}_{bB}^p (-1)^{pq} \mathbf{e}_B^q \\ \dot{\hat{\mathbf{Q}}}_b &= -\hat{\mathbf{d}}_{bi}^q \hat{\mathbf{e}}_i^p - \hat{\mathbf{d}}_{bB}^q \hat{\mathbf{e}}_B^p. \end{aligned} \tag{3.24}$$

Applying the duality relations

$$\hat{\mathbf{d}}_{ii}^q = (-1)^p (\mathbf{d}_{ii}^p)^T, \qquad \hat{\mathbf{d}}_{ib}^q = (-1)^p (\mathbf{d}_{bi}^p)^T, \qquad \hat{\mathbf{d}}_{bi}^q = (-1)^p (\mathbf{d}_{ib}^p)^T, \tag{3.25}$$

as well as

$$\mathbf{d}_{bB}^p = \mathbf{I}, \qquad \hat{\mathbf{d}}_{bB}^q = (-1)^{p-q} \mathbf{I}, \tag{3.26}$$

we can write

$$\frac{d}{dt} \begin{bmatrix} \mathbf{P}_i \\ \hat{\mathbf{Q}}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{d}_{ii}^p \\ -(\mathbf{d}_{ii}^p)^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} (-1)^p \hat{\mathbf{e}}_i^p \\ (-1)^{pq} \mathbf{e}_i^q \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{d}_{ib}^p \\ -(\mathbf{d}_{bi}^p)^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} (-1)^p \hat{\mathbf{e}}_b^p \\ (-1)^{pq} \mathbf{e}_b^q \end{bmatrix} \tag{3.27a}$$

and

$$\begin{bmatrix} (-1)^{pq} \mathbf{e}_B^q - \frac{d}{dt} \mathbf{P}_b \\ -(-1)^{p-q} \hat{\mathbf{e}}_B^p - \frac{d}{dt} \hat{\mathbf{Q}}_b \end{bmatrix} = \begin{bmatrix} \mathbf{0} & -\mathbf{d}_{bi}^p \\ (\mathbf{d}_{ib}^p)^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} (-1)^p \hat{\mathbf{e}}_i^p \\ (-1)^{pq} \mathbf{e}_i^q \end{bmatrix}. \tag{3.27b}$$

We can now state the following main result of this chapter:

Theorem 3.1. The discrete formulation of a system of two conservation laws with $p = n \in \{1, 2, 3\}$, $q = 1$ on two staggered grids with a system boundary that gives rise to the definition of a primal and a dual n-complex as sketched above, reads

$$\underbrace{\frac{d}{dt} \begin{bmatrix} \mathbf{P}_i \\ \hat{\mathbf{Q}}_i \end{bmatrix}}_{\dot{\mathbf{x}}} = \underbrace{(-1)^n \begin{bmatrix} \mathbf{0} & \mathbf{d}_{ii}^n \\ -(\mathbf{d}_{ii}^n)^T & \mathbf{0} \end{bmatrix}}_{\mathbf{J}} \underbrace{\begin{bmatrix} \hat{\mathbf{e}}_i^p \\ \mathbf{e}_i^q \end{bmatrix}}_{\mathbf{e}} + \underbrace{(-1)^n \begin{bmatrix} \mathbf{0} & \mathbf{d}_{ib}^n \\ -(\mathbf{d}_{bi}^n)^T & \mathbf{0} \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \hat{\mathbf{e}}_b^p \\ \mathbf{e}_b^q \end{bmatrix}}_{\mathbf{u}} \tag{3.28a}$$

$$\underbrace{(-1)^n \begin{bmatrix} \mathbf{e}_B^q \\ \hat{\mathbf{e}}_B^p \end{bmatrix} - \frac{d}{dt} \begin{bmatrix} \mathbf{P}_b \\ \hat{\mathbf{Q}}_b \end{bmatrix}}_{\mathbf{y}} = \underbrace{(-1)^n \begin{bmatrix} \mathbf{0} & -\mathbf{d}_{bi}^n \\ (\mathbf{d}_{ib}^n)^T & \mathbf{0} \end{bmatrix}}_{\mathbf{G}^T} \underbrace{\begin{bmatrix} \hat{\mathbf{e}}_i^p \\ \mathbf{e}_i^q \end{bmatrix}}_{\mathbf{e}}. \tag{3.28b}$$

$\mathbf{d}_{ii}^n$, $\mathbf{d}_{ib}^n$, $\mathbf{d}_{bi}^n$ are co-incidence matrices relating $(n-1)$-cells and n-cells on the primal complex. $\mathbf{P}_{i/b}$, $\hat{\mathbf{Q}}_{i/b}$, $\mathbf{e}_{i/b/B}^q$, $\hat{\mathbf{e}}_{i/b/B}^p$, are vector representations of the j-cochains with the integral values of the n-forms, 1-forms, $(n-1)$-forms and 0-forms on the corresponding discrete objects (j-chains) of the primal and the dual complex.

Proof. Primal and dual n-complex can be constructed in analogy to above for $n = 1$ and $n = 3$ with complete duality between i and b indexed cells and the definition of B indexed cells. This allows to apply the duality formulas (3.19), (3.20) to (3.27a), (3.27b), which yields (3.28a), (3.28b). □

The discrete formulation of the system of two conservation laws is *exact*. It is written in form of the input-output representation (2.13) of a finite-dimensional Dirac structure, for which the power balance $-\mathbf{e}^T\dot{\mathbf{x}} + \mathbf{y}^T\mathbf{u} = 0$ holds. However, it can not be understood as a finite-dimensional PH system, as the vector of co-energy variables $\mathbf{e}$ is *not* derived from a finite-dimensional energy function. A finite-dimensional PH system is obtained if the energy functional is replaced by a finite-dimensional approximation and discrete constitutive relations are established. In other words, or more general, the true boundary fluxes at the integration domains have to be replaced by *numerical flux functions*, which is the key ingredient of classical finite volume discretization.

Remark 3.5. The *collocated* pairs of boundary variables $(\hat{\mathbf{e}}_b^p, \mathbf{e}_B^q)$ and $(\mathbf{e}_b^q, \hat{\mathbf{e}}_B^p)$ are *not exactly power-conjugated*. This is due to the presence of $\frac{d}{dt}\mathbf{P}_b$ and $\frac{d}{dt}\hat{\mathbf{Q}}_b$ in the output equation (3.28b)[14]. *Vice versa*, $\mathbf{y}$ does not exactly represent fluxes at the system boundary. This effect, which is due to the use of staggered dual grids, decreases with grid refinement.

Remark 3.6. The results of [174] with uniform boundary inputs can be recovered if the system boundary is drawn exclusively along (n–1)-cells of the primal *or* the dual mesh.

3.4 Numerical Approximation

The topological information, coded in the primal and dual n-complex yields the *exact* discrete formulation of the two conservation laws in input-/output form (3.28a), (3.28b). The elements of the state vector $\mathbf{x}$ and the co-state/effort vector $\mathbf{e}$ are the *integral quantities* on the integration domains of the primal

[14] In [174], this fact is less obvious.

and the dual mesh. However, the constitutive relations (3.3) are formulated *locally* between the corresponding differential forms.

For a *numerical approximation* model in PH form, a *discrete energy* must be defined in terms of the discrete states, and the discrete efforts must be derived from this approximate energy. We present this *structure-preserving* discretization of the constitutive equations on the example of the 2D irrotational shallow water equations. We use a finite volumes approximation to compute the numerical fluxes and to define a lumped Hamiltonian, before we discuss the properties of our approach, in particular the relation of some aspects to "classical" finite volume schemes.

Remark 3.7. In [174], a *quadratic* discrete energy is directly expressed in terms of the cochains on both complexes. The *linear* constitutive relations are expressed involving the *discrete Hodge operator*[15]. Such a *direct formulation* of the approximate energy is less obvious for non-quadratic energies and spatial dependencies.

3.4.1 Example: Irrotational 2D Shallow Water Equations

Recall the 2D shallow water equations (SWE) in vector calculus notation[16], as presented in Subsection 2.3.2:

$$\begin{bmatrix} \partial_t h \\ \partial_t \mathbf{u} + \zeta \mathbf{u}^\perp \end{bmatrix} = \begin{bmatrix} 0 & -\mathrm{div} \\ -\mathrm{grad} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + gh + gb \\ h\mathbf{u} \end{bmatrix}. \tag{3.29}$$

$h(x,y)$ denotes the elevation of the free water surface over the bottom profile $b(x,y)$, $\mathbf{u}(x,y) = [u(x,y)\ \ v(x,y)]^T$ is the 2-dimensional velocity vector field and g the gravitational acceleration. The term $\zeta\mathbf{u}^\perp$ with $\mathbf{u}^\perp = [v\ \ -u]^T$ represents the acceleration due to rotation of the flow and stems from the transport term in the momentum equation. $\zeta = \partial_x v - \partial_y u$ denotes the *vorticity*. The vector of effort variables on the right of (3.29) contains the hydrodynamic pressure $p_{dyn} = \frac{1}{2}\mathbf{u}\cdot\mathbf{u} + gh + gb$ and the vector of discharge per unit width $h\mathbf{u}$ in x- and y-direction. p_{dyn} and the components hu and hv of the discharge vector field can be expressed as variational derivatives of the Hamiltonian H (equivalently partial derivatives of the Hamiltonian density $\mathcal{H}$) with respect to h, u and v with the energy per unit mass

$$H = \int_\Omega \mathcal{H}\, dxdy, \qquad \mathcal{H} = \frac{1}{2}h\mathbf{u}\cdot\mathbf{u} + \frac{1}{2}gh^2 + ghb. \tag{3.30}$$

We consider a rectangular domain $\Omega \subset \mathbb{R}^2$. For flow problems with negligible rotational acceleration $\zeta\mathbf{u}^\perp$ (e. g. a unidirectional flow in a flat bed), the *irrotational* SWE have the canonical form of a PH system according to Definition 2.4 with $n = p = 2$, $q = 1$. The states and co-states in terms of differential forms are then $p = *h$, $q = \mathbf{u}^\flat$, $e^p = p_{dyn}$ and $e^q = -*(h\mathbf{u})^\flat$. For brevity, we assume $b(x,y) \equiv 0$.

[15] See [47], Definition 6.1.
[16] See e. g. [5].

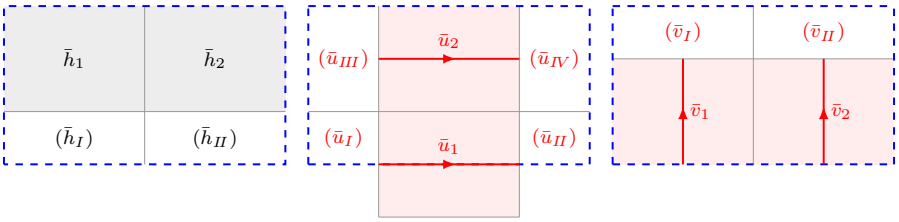

Figure 3.6: 2D (sub-)domains related to the discrete states.

3.4.2 Finite-Dimensional Port-Hamiltonian Model

We consider uniform rectangular grids that are shifted by half the grid size $\Delta x/2$ and $\Delta y/2$ in each direction. The *discrete state vector* (or integral state vector)

$$\mathbf{x}_d = \begin{bmatrix} \mathbf{h}_d^T & \mathbf{u}_d^T & \mathbf{v}_d^T \end{bmatrix}^T \tag{3.31}$$

with components[17] $h_{d,j}$, $j = 1, \ldots, |\mathcal{F}_i|$, $u_{d,k}$, $k = 1, \ldots, |\hat{\mathcal{E}}_i^x|$, and $v_{d,l}$, $l = 1, \ldots, |\hat{\mathcal{E}}_i^y|$, represents the approximate area integrals of $h(x, y)$ on the interior primal faces and the approximate line integrals of $u(x, y)$ and $v(x, y)$ on the horizontal/vertical interior edges of the dual grid. Denote

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{\mathbf{h}}^T & \bar{\mathbf{u}}^T & \bar{\mathbf{v}}^T \end{bmatrix}^T \tag{3.32}$$

the vector of *average* states on the corresponding primal 2-cells and dual 1-cells, which are given by

$$\bar{h}_j = \frac{h_{d,j}}{\Delta x \Delta y}, \qquad \bar{u}_k = \frac{u_{d,k}}{\Delta x}, \qquad \bar{v}_l = \frac{v_{d,l}}{\Delta y}, \tag{3.33}$$

where $\Delta x \Delta y = |f_j|$ is the area of a primal face and $\Delta x = |\hat{e}_k^x|$, $\Delta y = |\hat{e}_l^y|$ are the lengths of the dual edges. We understand the average state values as approximations of $h(x, y)$, $u(x, y)$ and $v(x, y)$ on the interior primal faces and the surrounding areas of the dual interior edges, respectively. These domains (see the shaded regions in Fig. 3.6) may lie partially outside Ω, which is the case if an interior dual edge lies on the system boundary. The superposition of the dual grids divides the whole spatial domain into *control volumes* with identical values of average states, indexed $I = (j, k, l) \in \mathcal{I}$, see Fig. 3.7. Their sizes for regular, uniformly shifted grids is $\Delta x \Delta y/4$. $\mathcal{I}$ denotes the set of all multi-indices on the superposition of the primal and the dual mesh.

We define the *discrete Hamiltonian* as

$$H_d(\mathbf{x}_d) = \frac{\Delta x \Delta y}{4} \sum_{I \in \mathcal{I}} \mathcal{H}\left(\frac{h_{d,j(I)}}{\Delta x \Delta y}, \frac{u_{d,k(I)}}{\Delta x}, \frac{v_{d,l(I)}}{\Delta y} \right), \tag{3.34}$$

[17] In the sequel, we use the indices j, k and l to refer to these components.

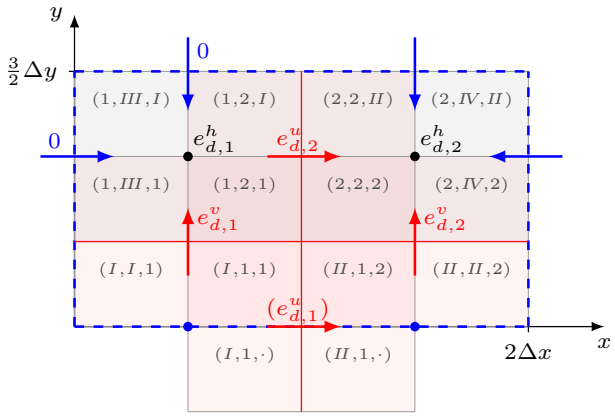

Figure 3.7: 2D control volumes with multi-index $I = (j, k, l)$ on a sample grid over $\Omega \in (0, 2\Delta x) \times (0, \frac{3\Delta y}{2}) \subset \mathbb{R}^2$. A dot "$\cdot$" denotes that the corresponding state is not needed to compute an effort. Black and red: Discrete efforts computed from $\nabla H(\mathbf{x}_d)$. As it does not affect a state differential equation, $e^u_{d,1}$ is set in parentheses. Blue: Boundary efforts = input variables.

where $j(I)$, $k(I)$, $l(I)$ are the components of the multi-index I. In the boundary regions, where no discrete states are defined, we need to impose additional *ghost values* for the states, denoted in brackets in Fig. 3.6. We assign *constant* ghost values, based on reasonable assumptions, e. g. given boundary conditions or the steady state. The consistency of the effort approximation depends on the validity of these assumptions, see further below.

Remark 3.8. Usually, the ghost values are computed by extrapolation from the interior discrete states[18]. We could do accordingly, and assign the adjacent discrete states to the ghost cells (zero order extrapolation). For a consistent energy approximation, this could (as a possible interpretation) impose an enlargement of the j-cells to which the discrete states are associated. This re-interpretation of the discrete states and their spatial domains is, however, not consistent with the exact PH representation of the conservation laws (3.28a), (3.28b).

The vector of *discrete efforts*

$$\mathbf{e}_d = \left[(\mathbf{e}_d^h)^T \quad (\mathbf{e}_d^u)^T \quad (\mathbf{e}_d^v)^T\right]^T \tag{3.35}$$

is derived from the discrete Hamiltonian:

$$\mathbf{e}_d := \nabla H_d(\mathbf{x}_d). \tag{3.36}$$

[18] See e. g. [115], Chapter 7.

In particular, we express the single discrete efforts as

$$\begin{aligned} e^h_{d,j} &:= \frac{\partial H_d}{\partial h_{d,j}} = \sum_{I\in\mathcal{I}^s_j} \frac{\Delta x\Delta y}{4}\frac{1}{\Delta x\Delta y}\left.\frac{\partial \mathcal{H}}{\partial h}\right|_{\bar{\mathbf{x}}_I}, \\ e^u_{d,k} &:= \frac{\partial H_d}{\partial u_{d,k}} = \sum_{I\in\mathcal{I}^u_k} \frac{\Delta x\Delta y}{4}\frac{1}{\Delta x}\left.\frac{\partial \mathcal{H}}{\partial u}\right|_{\bar{\mathbf{x}}_I}, \\ e^v_{d,l} &:= \frac{\partial H_d}{\partial v_{d,l}} = \sum_{I\in\mathcal{I}^v_l} \frac{\Delta x\Delta y}{4}\frac{1}{\Delta y}\left.\frac{\partial \mathcal{H}}{\partial v}\right|_{\bar{\mathbf{x}}_I}. \end{aligned} \tag{3.37}$$

The notation $(\cdot)|_{\bar{\mathbf{x}}_I}$ denotes the evaluation of the partial derivatives of $\mathcal{H}$ at $(\bar{h}_j, \bar{u}_k, \bar{v}_l) = (\frac{h_{d,j(I)}}{\Delta x\Delta y}, \frac{u_{d,k(I)}}{\Delta x}, \frac{v_{d,l(I)}}{\Delta y})$. $\mathcal{I}^h_j$, $\mathcal{I}^u_k$ and $\mathcal{I}^v_l$ are the sets of multi-indices that refer to the 2×2 control volumes associated with $h_{d,j}$, $u_{d,k}$ or $v_{d,l}$, respectively. For the corresponding average efforts

$$\bar{\mathbf{e}} = \begin{bmatrix} (\mathbf{e}^h_d)^T & \frac{(\mathbf{e}^u_d)^T}{\Delta y} & \frac{(\mathbf{e}^v_d)^T}{\Delta x} \end{bmatrix}^T, \tag{3.38}$$

we obtain

$$\bar{e}^h_j = \sum_{I\in\mathcal{I}^h_j} \frac{1}{4}\left.\frac{\partial \mathcal{H}}{\partial h}\right|_{\bar{\mathbf{x}}_I}, \qquad \bar{e}^u_k = \sum_{I\in\mathcal{I}^u_k} \frac{1}{4}\left.\frac{\partial \mathcal{H}}{\partial u}\right|_{\bar{\mathbf{x}}_I}, \qquad \bar{e}^v_l = \sum_{I\in\mathcal{I}^v_l} \frac{1}{4}\left.\frac{\partial \mathcal{H}}{\partial v}\right|_{\bar{\mathbf{x}}_I}. \tag{3.39}$$

Theorem 3.2. The finite-dimensional PH model

$$\dot{\mathbf{x}}_d = \mathbf{J}\mathbf{e}_d + \mathbf{G}\mathbf{e}_b \tag{3.40a}$$

$$\mathbf{y}_d = \mathbf{G}^T\mathbf{e}_d, \tag{3.40b}$$

with $\mathbf{x}_d$ the discrete state vector (3.31), the effort vector $\mathbf{e}_d = \nabla H_d(\mathbf{x}_d)$ derived from the discrete Hamiltonian (3.34), the boundary input vector $\mathbf{e}_b = \begin{bmatrix}(\hat{\mathbf{e}}^p_b)^T & (\mathbf{e}^q_b)^T\end{bmatrix}^T$ and the matrices $\mathbf{J}$ and $\mathbf{G}$ as defined in (3.28), is a consistent approximation of the irrotational 2D shallow water equations, if the ghost values on the boundary control volumes are consistent with the boundary conditions.

Proof. We consider the differential equation of the finite-dimensional PH model in terms of the *average* states, efforts and boundary inputs (see Appendix B.1.1):

$$\dot{\bar{\mathbf{x}}} = \frac{1}{\Delta x}\bar{\mathbf{J}}\bar{\mathbf{e}} + \frac{1}{\Delta x}\bar{\mathbf{G}}\bar{\mathbf{e}}_b, \tag{3.41}$$

where the elements of $\bar{\mathbf{J}}$, $\bar{\mathbf{G}}$ are from the set $\{0, \pm 1, \pm\frac{\Delta y}{\Delta x}\}$. The (local) approximation error, which determines the consistency order[19], is

$$\epsilon_{loc,\Delta x} = \left\| \dot{\bar{\mathbf{x}}}^* - \frac{1}{\Delta x}\bar{\mathbf{J}}\,\bar{\mathbf{e}}|_* - \frac{1}{\Delta x}\bar{\mathbf{G}}\,\bar{\mathbf{e}}_b \right\|_{\Delta x} \tag{3.42}$$

[19] See [88], Eq. (16.19).

for $\Delta x \to 0$. $\dot{\bar{\mathbf{x}}}^*$ contains the time derivatives of the *exact solution* at the centers of the 2×2 control volumes, and is given by the right hand side of the PDE (3.29). The average boundary inputs $\bar{\mathbf{e}}_b$ are exactly known. The average efforts $\bar{\mathbf{e}}$ are computed according to the finite volumes approximation, but assuming that the underlying distribution of the states solves the PDE. The $*$ is a shortcut for the substitution

$$* = \left(\bar{h}_j = \int_{y^h_{j,\mathrm{lo}}}^{y^h_{j,\mathrm{up}}} \int_{x^h_{j,\mathrm{le}}}^{x^h_{j,\mathrm{ri}}} \frac{h(x,y)}{\Delta x \Delta y}\, dx dy, \right.$$
$$\left. \bar{u}_k = \int_{x^u_{k,\mathrm{le}}}^{x^u_{k,\mathrm{ri}}} \frac{u(x, y^u_k)}{\Delta x}\, dx, \quad \bar{v}_l = \int_{y^v_{l,\mathrm{lo}}}^{y^v_{l,\mathrm{up}}} \frac{v(x^v_l, y)}{\Delta y}\, dy \right). \quad (3.43)$$

The limits of integration with the subscripts *le*, *ri*, *lo*, *up* refer to the left, right, lower and upper boundaries of the considered 2×2 control volume in x- and y-direction. y^u_k and x^v_l denote the corresponding center coordinates. The norm

$$\|\mathbf{f}\|_{\Delta x} := \left(\Delta x \sum_{j=1}^{N} |f_j|^2 \right)^{\frac{1}{2}} \quad (3.44)$$

is the discrete counterpart of the L^2-norm for functions. The boundary inputs being exactly known, they cancel from (3.42). The order of the error $\epsilon_{loc,\Delta x}$ is certainly $\mathcal{O}(\Delta x^p)$, (with an integer p) if for all indices j, k, and l, the errors

$$\begin{aligned} \epsilon^{h,x}_j &= \left| \frac{\partial}{\partial x} e^h_j - \frac{1}{\Delta x}\, \bar{e}^h_j|_* \right|, & \epsilon^{h,y}_j &= \left| \frac{\partial}{\partial y} e^h_j - \frac{1}{\Delta y}\, \bar{e}^h_j|_* \right| \\ \epsilon^u_k &= \left| \frac{\partial}{\partial x} e^u_k - \frac{1}{\Delta x}\, \bar{e}^u_k|_* \right|, & \epsilon^l_v &= \left| \frac{\partial}{\partial y} e^v_l - \frac{1}{\Delta y}\, \bar{e}^v_l|_* \right| \end{aligned} \quad (3.45)$$

are of order $\mathcal{O}(\Delta x^p)$ for $\Delta x, \Delta y \to 0$. We denote e^h_j, e^u_k, e^v_l the *true* efforts at the centers of the 2×2 control volumes. By Taylor series expansion, it can be verified that on superposed, rectangular grids with constant grid size $\frac{\Delta x}{2} \times \frac{\Delta y}{2}$, the errors (3.45) are of order $\mathcal{O}(\Delta x^2)$, if their computation does *not* involve ghost values for the average states. If the shifts between primal and dual grid are different from $\frac{\Delta x}{2}$, $\frac{\Delta y}{2}$, the order decreases to $\mathcal{O}(\Delta x)$. If the discrete efforts are computed based on ghost values, the consistency error is of order ≥ 1 only if the ghost values are consistent with the boundary conditions. In Appendix B.1.2, the computations of the consistency errors for the efforts depicted in Fig. 3.7 are sketched: based on (i) no, (ii) consistent, and (iii) inconsistent ghost values. □

3.4.3 Remarks

The numerical approximation of the Hamiltonian and the efforts, and the subsequent consistency analysis, give rise to the following complementary remarks.

1. The output equation of the discretized average model is simply written $\bar{\mathbf{y}} = \mathbf{G}^T\bar{\mathbf{e}}$, without scaling. Assigning constant ghost values for the states on the boundary integration domains corresponds to $\frac{d}{dt}\mathbf{P}_b = \mathbf{0}$, $\frac{d}{dt}\hat{\mathbf{Q}}_b = \mathbf{0}$ in (3.28b). Consequently, $\mathbf{y}_d$ and $\bar{\mathbf{y}}$ can be understood as integral/average numerical approximations of the output boundary efforts for $\Delta x, \Delta y \to 0$.

2. The resulting finite-dimensional model is in PH form, i. e. the discretization scheme is *structure-preserving*. The PH structure implies, for H_d positive definite, Lyapunov stability of the unforced equilibrium.

3. For the average discretized model in PH form, *numerical stability* (more precisely, numerical stability of the semi-discretization method, see [88], Section 16.2) can be shown: For bounded $\bar{\mathbf{e}}_b(t)$, there exists on every time interval $[0, t^*]$ a bound $c(t^*) < \infty$ such that $\|\bar{\mathbf{x}}(t)\|_{\Delta x} < c(t^*)$.

4. As discussed above, the consistency order of the effort approximation for the nonlinear SWE is 2 inside the spatial domain. This is due to the uniform shift of the primal and the dual grid, which implies a centered approximation of the constitutive equations. The order can be increased by computing the numerical fluxes based on a wider stencil, using a semi-discrete *generalized Leapfrog* scheme, see [87], or [64] from the finite-difference perspective. This approach was applied in [95] for the finite-volume approximation of 1D hyperbolic PH systems and extended in [173] to the 2D case.

5. The grid shifts in x- and y-direction can be understood as design degrees of freedom to parametrize numerical schemes which take into account the direction of propagation (of the solution), in the sense of *upwinding*[20]. For such non-centered schemes, the consistency order inside the spatial domain reduces to 1. For the upwinding interpretation of structure-preserving discretization schemes based on mixed finite elements, see Chapter 4.

6. The bottleneck for the consistency order of the overall numerical scheme is the assignment of *constant* ghost values, which can be inconsistent with the boundary conditions, e. g. at an outflow boundary. The usual approach to extrapolate the ghost values from the interior discrete states, ensures consistency. This measure, however, disturbs the PH structure of the approximate model, as not all numerical fluxes are derived exclusively from the discrete, time-invariant Hamiltonian. The corresponding numerical error acts as a disturbance to the PH model. Its effect can be dissipated if the model contains physical damping.

[20] We write *in the sense of*, as the grid shift is fixed and based on *a priori* assumptions on the flow direction. For upwinding methods, see e. g. [115], Chapter 4.

7. If the Hamiltonian is separable, as in the example of an acoustic duct [186], the computation of the discrete efforts does not rely on the ghost values, which guarantees consistency of the scheme. Note also the recently proposed structure-preserving finite-difference scheme for hyperbolic PH systems, which has been applied to the above mentioned example [187], [188].

8. For linear PDEs, consistency and numerical stability of the approximation directly imply convergence of the numerical scheme according to the *Lax-Richtmyer equivalence theorem*[21]. For a further discussion on convergence, we refer to the corresponding literature, e. g. [158], [52], [115], or more specifically, [53][22].

3.5 Conclusions

We proposed the exact integral/discrete PH formulation of hyperbolic systems of two conservation laws on staggered grids on n-dimensional spatial domains, by using the topological information in terms of (co-)incidence matrices of the related n-complexes. We extended known results by allowing for boundary input variables of mixed type as a basis to tackle a wider class of control problems with a non-uniform causality of the boundary port. We performed the approximation of the energy to obtain a numerical (simulation) model with the classical finite volume approach. On the nonlinear example of the 2D irrotational shallow water equations, we showed and discussed the consistency of the approximation in the domain and on the boundary.

The exact integral representation of the conservation laws by a finite set of differential equations can serve as a basis for observer design [100] and feedback control. In terms of trajectory planning/feedforward control, the presented discretization in space can be combined with symplectic time integration. To obtain a discrete-time flat parametrization of states and inputs, as presented in Section 6.4 for the 1D case, special care must be taken about the consistent approximation of the constitutive equations in two spatial dimensions. Another interesting question is to understand the shifts between primal and dual grid as design parameters for a control-oriented model in order to account for the natures of different systems of conservation laws, e. g. their "ratio" between convection and diffusion. This question is related to the discussions in Chapter 4, where a mixed FEM approach is applied both to hyperbolic and parabolic systems.

Currently, the presented approach is applied to model the heat and mass transfer through catalytic foams with an irregular structure [168] (see also the INFIDHEM[23] project website [75]). For the heat transfer on the solid and the

[21] See [88], Section 13.2 for semi-discretization.

[22] As a related steady-state problem, the convergence of a centered scheme on two staggered finite volume grids for the incompressible Navier-Stokes equations in 2D is discussed.

[23] *Interconnected Infinite-Dimensional Systems for Heterogeneous Media*, cofunded by *Deutsche Forschungsgemeinschaft* and *Agence Nationale de la Recherche*.

Figure 3.8: Pictures of a metallic foam with an irregular structure. Samples provided by Marie-Line Zanota, LGPC Lyon. Photos taken by Tobias Scheuermann, TUM Garching.

fluid phase, and the heat exchange between both of them, the material structure imposes a cell complex, on which temperature differences (discrete driving forces) are evaluated. The internal energy is balanced on a dual complex, which is constructed as the *barycentric* dual. Instead of a straightforward finite volume approximation on cartesian grids, the irregularity of the dual complexes imposes the use of *interpolation functions* as proposed in *Tonti's cell method* [185]. Note that the approach presented in this chapter allows to easily incorporate boundary conditions both on the temperature (Dirichlet) and the heat flux (Neumann) in the explicit simulation and control model.

Chapter 4

Mixed Galerkin Discretization

In this chapter[1], we present an alternative method for the structure-preserving discretization of port-Hamiltonian systems of conservation laws. Unlike the approach described in the preceding section, the procedure starts on a *single mesh*, over which the mixed approximation spaces for the power variables are defined. More general, the dual character of physical variables is not accounted for in a first step, in the sense that differential forms of the same degree, but different (inner or outer) orientation are treated identically. A *mixed Galerkin* approximation yields a linear system of equations that relates the vectors of discrete power variables. To obtain a non-degenerate discrete power balance (as a property of the finite-dimensional Dirac structure), based on which an approximate PH model can be defined, appropriate *power-preserving* mappings of the discrete power variables are necessary. The mappings of the original degrees of freedom will allow for an interpretation of the *minimal* discrete flows (or states) and efforts in terms of *topological duality*. This interpretation is exploited in the consistent approximation of the constitutive equations[2]. The parameters in the power-preserving mappings can be used to adapt the approximation to the nature of the underlying system (hyperbolic vs. parabolic).

We present the geometric discretization of distributed parameter port-Hamiltonian (dPH) systems based on the *weak formulation* of the underlying Stokes-Dirac structure. Doing so, some limitations and restrictions of current approaches can be overcome.

- The strict separation of *metric-independent structure* and *constitutive equations* is maintained in our approach.
- Our formulation is valid for systems on spatial domains with *arbitrary dimension*.

[1] The chapter is mainly based on the article [104]. In addition, it contains results on the approximation of the 1D wave and the 1D heat equation as presented in [97] and [96].

[2] Recall that the Hodge star in the local representation of the constitutive equation expresses this duality.

- Boundary inputs[3] are imposed weakly, i. e. they *appear directly* in the weak formulation of the Stokes-Dirac structure and the finite-dimensional approximation.
- The power-preserving maps for the discrete bond variables offer design degrees of freedom to parametrize the resulting finite-dimensional PH state space models. They allow for trade-offs between a centered approximation and upwinding, according to the nature of the considered system.
- Mapping the flow variables instead of the efforts *avoids a structural artificial feedthrough*, which is not desirable for the approximation of hyperbolic systems.

We consider as the prototypical example of dPH systems, an *open system of two hyperbolic conservation laws in canonical form*, as introduced in Section 2.2. The language of differential forms, see Appendix A.1, emphasizes the geometric nature of each variable and allows for a unifying representation independent of the dimension of the spatial domain.

The chapter is structured as follows. Based on the representation of a Stokes-Dirac structure with non-uniform boundary causality according to Theorem 2.4, we propose the *weak form of the Stokes-Dirac structure* in Section 4.1. Section 4.2 deals with the *mixed Galerkin approximation* of this Stokes-Dirac structure. Due to the different geometric nature of the power variables and their approximation spaces, the discrete power balance involves *degenerate* duality pairings. In Section 4.3, we define *minimal* discrete power variables (pairs of bond variables) with *non-degenerate* duality products by *power-preserving mappings*. The so-defined subspace of the bond space is a *Dirac structure* which admits different representations. The *explicit* input-output representation, together with the finite-dimensional approximation of the Hamiltonian, leads to the desired PH approximate models in state space form, which are given in Section 4.4. In Section 4.5, we recall *Whitney forms*, which shall be used as geometric finite elements in the sequel. We apply our discretization method to the 1D wave and the 1D heat equation in Section 4.6 and illustrate the effects of the discretization parameter on the approximation of the constitutive equations. For both examples, we analyze the model structures and assess optimal parameter choices based on the solution of initial value problems. We compare the eigenvalue approximation of the 1D wave equation with the method of [71]. Certain parameter choices can be interpreted in terms of *upwinding*, which is particularly favorable for hyperbolic systems. For the parabolic heat equation, the analysis of zeros and eigenvalues shows the superiority of a *centered* approximation. Section 4.7 is devoted to the illustration of our method on the example of the 2D wave equation on a rectangular simplicial mesh. We derive the power-preserving mappings based on elementary examples and we emphasize the interpretation of the finite-dimensional state and power variables in

[3] In-domain inputs can be treated identically.

terms of integral quantities on the grid. The dependence of the approximation quality on the mapping parameters is illustrated with a 2D simulation study, which again reveals the favorable effect of an upwinding parameter choice. The simulation of the double slit experiment shows the applicability of our method to a problem on a non-trivial geometry. Section 4.8 closes the chapter with a summary and comments on ongoing work.

4.1 Weak Form of the Stokes-Dirac Structure

The first motivation to study the approximation of dPH systems based on their *weak* form is the fact that most of the common numerical methods in engineering, including commercial tools, are based on a Galerkin-type finite-dimensional approximation of the PDEs in weak form[4]. Also in the context of existing works on linear dPH systems in one spatial dimension, this perspective is natural. The statements on well-posedness and stability based on the theory of C_0 semigroups rely on the *mild* solution of the abstract (operator) differential equation. These solutions, however, corresponds to the weak solutions, as known from the theory of PDEs, see [89], page 127: *"In fact, the concept of a mild solution is the same as the concept of a weak solution used in the study of partial differential equations."* A third point, which motivates to discretize dPH systems based on their weak form, is the close relation with *discrete exterior calculus* (i. e. the mathematical formalism for integral modeling of conservation laws), which has been used in [174] for PH systems: "*Note that the process of integration to suppress discontinuity is, in spirit, equivalent to the idea of weak form used in the Finite Element method*" [48]. Finally, also in Bossavit's work on the mixed geometric discretization for computational electromagnetism [20], [21], the quality of a *weak* formulation is addressed "*How weak is the weak solution in finite element methods?*" [22].

The *weak form* of the Stokes-Dirac structure of Theorem 2.4 is obtained by a *duality pairing* (which involves the exterior product and integration) on Ω with *test forms* of appropriate degrees which do *not* vanish on the boundary[5]. The latter allows for a *weak* imposition of the input boundary conditions $u_i^q = e_i^\Gamma$, $i = 1, \ldots, n_\Gamma$, and $u_j^p = \hat{e}_j^\Gamma$, $j = 1, \ldots, \hat{n}_\Gamma$.

[4] We use the *weak form* and not the *variational form*. The reason is that we focus on the geometric structure of the equations and do not mention the associated variational problem. We refer to [213] and [200] for the link of the variational problem in Lagrangian mechanics in finite and infinite dimension with a Dirac structure. Note that this link is less obvious e. g. for non-Hamiltonian fluids, which are described by a non-canonical structure, see e. g. [137], [32].

[5] In the weak formulation of boundary value problems, mostly test functions with compact support inside Ω are chosen such that boundary conditions have to be imposed *directly* on the solution. This is however not mandatory. By test functions which are non-zero on $\partial\Omega$, boundary conditions can be imposed in a *weak* fashion, cf. [158], Section 14.3.1, p. 483.

Definition 4.1. The weak form of the Stokes-Dirac structure of Theorem 2.4 is given by the subspace $\mathcal{D} \subset \mathcal{F} \times \mathcal{E}$ with $\mathcal{F}$ and $\mathcal{E}$ as in (2.30), (2.31), where

$$\langle v^p | f^p \rangle_\Omega = \langle v^p | (-1)^r \mathrm{d} e^q \rangle_\Omega \qquad \forall v^p \in H^1\Lambda^{n-p}(\Omega), \tag{4.1a}$$

$$\langle v^q | f^q \rangle_\Omega = \langle v^q | \mathrm{d} e^p \rangle_\Omega \qquad \forall v^q \in H^1\Lambda^{n-q}(\Omega) \tag{4.1b}$$

holds and the boundary port variables are defined by (2.29).

Applying integration by parts according to (A.7), we obtain the weak form of the Stokes-Dirac structure with *weak treatment* of the boundary port variables.

Proposition 4.1. The *weak form* of the Stokes-Dirac structure in Theorem 2.4 with *weak treatment* of the boundary port variables is given by the subset $\mathcal{D} \subset \mathcal{F} \times \mathcal{E}$, $\mathcal{F}$ and $\mathcal{E}$ as in (2.30), where

$$\begin{aligned} \langle v^p | f^p \rangle_\Omega =& (-1)^{r+q} \langle \mathrm{d} v^p | e^q \rangle_\Omega \\ &- (-1)^{r+q} \sum_{i=1}^{n_\Gamma} \langle \operatorname{tr} v^p | e_i^\Gamma \rangle_{\Gamma_i} - (-1)^{r+q} \sum_{j=1}^{\hat{n}_\Gamma} \langle \operatorname{tr} v^p | \hat{f}_j^\Gamma \rangle_{\hat{\Gamma}_j} \end{aligned} \tag{4.2a}$$

$$\begin{aligned} \langle v^q | f^q \rangle_\Omega =& (-1)^{p} \langle \mathrm{d} v^q | e^p \rangle_\Omega \\ &- (-1)^{p} \sum_{i=1}^{n_\Gamma} \langle \operatorname{tr} v^q | f_i^\Gamma \rangle_{\Gamma_i} \qquad - (-1)^{p} \sum_{j=1}^{\hat{n}_\Gamma} \langle \operatorname{tr} v^q | \hat{e}_j^\Gamma \rangle_{\hat{\Gamma}_j} \end{aligned} \tag{4.2b}$$

holds for all test forms $v^p \in H^1\Lambda^{n-p}(\Omega)$ and $v^q \in H^1\Lambda^{n-q}(\Omega)$.

Proof. Equations (4.2) follow from (4.1) via integration by parts and the identities

$$\begin{aligned} (-1)^p \langle \operatorname{tr} v^p | \operatorname{tr} e^q \rangle_{\partial\Omega} &= \sum_{i=1}^{n_\Gamma} \langle \operatorname{tr} v^p | e_i^\Gamma \rangle_{\Gamma_i} + \sum_{j=1}^{\hat{n}_\Gamma} \langle \operatorname{tr} v^p | \hat{f}_j^\Gamma \rangle_{\hat{\Gamma}_j}, \\ \langle \operatorname{tr} v^q | \operatorname{tr} e^p \rangle_{\partial\Omega} &= \sum_{i=1}^{n_\Gamma} \langle \operatorname{tr} v^q | f_i^\Gamma \rangle_{\Gamma_i} + \sum_{j=1}^{\hat{n}_\Gamma} \langle \operatorname{tr} v^q | \hat{e}_j^\Gamma \rangle_{\hat{\Gamma}_j}. \end{aligned} \tag{4.3}$$

The latter are due to the definition (2.29) of boundary port variables and the definition of the subsets Γ_i, $\hat{\Gamma}_j$, which cover $\partial\Omega$ and whose intersections have zero measure. □

Remark 4.1. The latter representation of the Stokes-Dirac structure – if considered on a single control volume – is suitable for *discontinuous Galerkin* schemes, see e. g. [81], where the boundary terms are replaced by suitable *numerical fluxes*.

Remark 4.2. Note that the two conservation laws are described by the canonical matrix differential operator in (2.32), which contains only exterior derivatives. The weak form of the Stokes-Dirac structure is defined based on the *metric-independent* duality product arising from the integration-by-parts formula (A.7), applied to *both* conservation laws in Eq. (2.32). This is a difference to other approaches like the mixed mimetic discretization of the Stokes flow in [105] or the structure-preserving PH discretization in [56], where integration by parts is only applied to the equations that contain the *metric-dependent* codifferential.

Using the effort forms as test forms, $v^p = e^p$, $v^q = e^q$, and adding both equations of (4.2), we obtain after some reformulations the structural power balance (2.33), i. e.

$$\langle e^p | f^p \rangle_\Omega + \langle e^q | f^q \rangle_\Omega + \sum_{i=1}^{n_\Gamma} \langle e_i^\Gamma | f_i^\Gamma \rangle_{\Gamma_i} + \sum_{j=1}^{\hat{n}_\Gamma} \langle \hat{f}_j^\Gamma | \hat{e}_j^\Gamma \rangle_{\hat{\Gamma}_j} = 0. \tag{4.4}$$

We have arrived at a weak representation of the Stokes-Dirac structure of Theorem 2.4, which suits to establish discretized mixed Galerkin models of PH systems of two conservation laws.

4.2 Approximation of the Stokes-Dirac Structure

We introduce the *mixed Galerkin* approximation of the weak form of the Stokes-Dirac structure for a system of two conservation laws. Mixed or *duality* methods have been introduced to include constraints like the divergence-freedom of flows or to take account for the precise approximation of additional physical variables in the numerical approximation, see [27] as a classical reference for mixed finite elements. The duality of the power variables in the Stokes-Dirac structure – illustrated by the degrees of the differential forms – imposes the use of a mixed approximation. Expressing (4.2) in appropriate subspaces, and defining *in- and output port variables* whose pairings represent the transmitted power over the boundary, we obtain a finite number of equations for the Galerkin coefficients. On the so-defined subset of the discrete bond space, a discrete power continuity equation holds. Due to the different dimensions of the geometrically chosen approximation spaces, the bilinear forms that define the discrete power pairings are, however, *degenerate*.

4.2.1 Weak Imposition of Boundary Conditions

The *boundary inputs* are weakly imposed as boundary conditions, and appear *immediately* in the finite-dimensional system of equations for the Galerkin degrees of freedom. *Boundary outputs* are constructed via the discrete power balance. This point of view, which leads to state space models in input-output

form, distinguishes the structure-preserving discretization of PH systems from classical approaches to the numerical approximation of PDEs.

For the compactness of notation, we omit to explicitly write out the trace operator on the subsets of the boundary, i. e. $\langle v^p | e^q \rangle_{\Gamma_i} := \langle \operatorname{tr} v^p | \operatorname{tr} e^q \rangle_{\Gamma_i}$ etc. in the sequel. We start with the representation[6]

$$\begin{aligned}\langle v^p | f^p \rangle_\Omega =& (-1)^{r+q} \langle \mathrm{d} v^p | e^q \rangle_\Omega \\ &- (-1)^{r+q} \sum_{\mu=1}^{n_\Gamma} \langle v^p | e^q \rangle_{\Gamma_\mu} - (-1)^{r+q} \sum_{\nu=1}^{\hat{n}_\Gamma} \langle v^p | e^q \rangle_{\hat{\Gamma}_\nu} \qquad (4.5a)\\ \langle v^q | f^q \rangle_\Omega =& (-1)^{p} \langle \mathrm{d} v^q | e^p \rangle_\Omega \\ &- (-1)^{p} \sum_{\mu=1}^{n_\Gamma} \langle v^q | e^p \rangle_{\Gamma_\mu} - (-1)^{p} \sum_{\nu=1}^{\hat{n}_\Gamma} \langle v^q | e^p \rangle_{\hat{\Gamma}_\nu}, \qquad (4.5b)\end{aligned}$$

i. e. (4.2) without the explicit denomination of the boundary port variables. For a *mixed Galerkin* approximation of the Stokes-Dirac structure, we

- use different (*dual* or *mixed*) bases to approximate the spaces of flow and effort forms and
- from these bases, we choose the appropriate ones to approximate the test forms (*Galerkin* method).

Taking the test forms from the effort bases is the most obvious choice for the approximation of the Stokes-Dirac structure, as the resulting (discrete) duality pairings have an immediate interpretation in terms of power, see Eq. (4.4).

4.2.2 Approximation Problem and Compatibility Condition

The flow differential forms will be approximated by linear combinations of the basis forms of the subspaces

$$\begin{array}{rcccl} \Psi_h^p & = & \operatorname{span}\{\psi_1^p, \ldots, \psi_{N_p}^p\} & \subset & L^2\Lambda^p(\Omega), \\ \Psi_h^q & = & \operatorname{span}\{\psi_1^p, \ldots, \psi_{N_q}^q\} & \subset & L^2\Lambda^q(\Omega). \end{array} \qquad (4.6)$$

The subspaces for the effort and test forms are, accordingly,

$$\begin{array}{rcccl} \Phi_h^p & = & \operatorname{span}\{\varphi_1^p, \ldots, \varphi_{M_p}^p\} & \subset & H^1\Lambda^{n-p}(\Omega), \\ \Phi_h^q & = & \operatorname{span}\{\varphi_1^q, \ldots, \varphi_{M_q}^q\} & \subset & H^1\Lambda^{n-q}(\Omega). \end{array} \qquad (4.7)$$

From the trace theorem for H^1 spaces (as discussed in Subsection A.1.3), we know that the extension of the latter spaces to the boundary is in L^2. The subscript $h > 0$ denotes the discretization parameter[7] and we assume an *appropriate* choice of approximation spaces, i. e. for a given functional space V and its

[6] In the sequel, we denote portions of the boundary with greek indices and elements of the approximation subspaces with latin indices.

[7] Which corresponds to the spatial extent of finite elements or the inverse of the polynomial approximation order.

approximation V_h (see [158], Section 5.2) it is true that $\inf_{v_h \in V_h} \|v - v_h\| \to 0$ for all $v \in V$ if $h \to 0$. The *mixed Galerkin approximation problem* is as follows: Find approximate flow and effort forms

$$\begin{aligned} f_h^p(\mathbf{z}) &= \sum_{k=1}^{N_p} f_k^p \psi_k^p(\mathbf{z}) = \langle \mathbf{f}^p | \boldsymbol{\psi}^p(\mathbf{z}) \rangle \in \Psi_h^p, \\ f_h^q(\mathbf{z}) &= \sum_{l=1}^{N_q} f_l^q \psi_l^q(\mathbf{z}) = \langle \mathbf{f}^q | \boldsymbol{\psi}^q(\mathbf{z}) \rangle \in \Psi_h^q, \end{aligned} \tag{4.8}$$

and

$$\begin{aligned} e_h^p(\mathbf{z}) &= \sum_{i=1}^{M_p} e_i^p \varphi_i^p(\mathbf{z}) = \langle \mathbf{e}^p | \boldsymbol{\varphi}^p(\mathbf{z}) \rangle \in \Phi_h^p, \\ e_h^q(\mathbf{z}) &= \sum_{j=1}^{M_q} e_j^q \varphi_j^q(\mathbf{z}) = \langle \mathbf{e}^q | \boldsymbol{\varphi}^q(\mathbf{z}) \rangle \in \Phi_h^q, \end{aligned} \tag{4.9}$$

where $\langle \cdot | \cdot \rangle$ denotes the standard inner product on $\mathbb{R}^n$ as in Definition 2.1, such that

$$\begin{aligned} \langle v_h^p | f_h^p \rangle_\Omega =& (-1)^{r+q} \langle \mathrm{d} v_h^p | e_h^q \rangle_\Omega \\ & - (-1)^{r+q} \sum_{\mu=1}^{n_\Gamma} \langle v_h^p | e_h^q \rangle_{\Gamma_\mu} - (-1)^{r+q} \sum_{\nu=1}^{\hat{n}_\Gamma} \langle v_h^p | e_h^q \rangle_{\hat{\Gamma}_\nu}, \end{aligned} \tag{4.10a}$$

$$\begin{aligned} \langle v_h^q | f_h^q \rangle_\Omega =& (-1)^{p} \langle \mathrm{d} v_h^q | e_h^p \rangle_\Omega \\ & - (-1)^{p} \sum_{\mu=1}^{n_\Gamma} \langle v_h^q | e_h^p \rangle_{\Gamma_\mu} - (-1)^{p} \sum_{\nu=1}^{\hat{n}_\Gamma} \langle v_h^q | e_h^p \rangle_{\hat{\Gamma}_\nu} \end{aligned} \tag{4.10b}$$

hold for all $v_h^p \in \Phi_h^p$, $v_h^q \in \Phi_h^q$. The *discrete flow and effort vectors*

$$\begin{aligned} \mathbf{f}^p &= [f_1^p, \ldots, f_{N_p}^p]^T, & \mathbf{e}^p &= [e_1^p, \ldots, e_{M_p}^p]^T, \\ \mathbf{f}^q &= [f_1^q, \ldots, f_{N_q}^q]^T, & \mathbf{e}^q &= [e_1^q, \ldots, e_{M_q}^q]^T \end{aligned} \tag{4.11}$$

contain the approximation coefficients, and the vectors (we omit the argument $\mathbf{z}$ in the sequel)

$$\begin{aligned} \boldsymbol{\psi}^p(\mathbf{z}) &= [\psi_1^p(\mathbf{z}), \ldots, \psi_{N_p}^p(\mathbf{z})]^T, & \boldsymbol{\varphi}^p(\mathbf{z}) &= [\varphi_1^p(\mathbf{z}), \ldots, \varphi_{M_p}^p(\mathbf{z})]^T, \\ \boldsymbol{\psi}^q(\mathbf{z}) &= [\psi_1^q(\mathbf{z}), \ldots, \psi_{N_q}^q(\mathbf{z})]^T, & \boldsymbol{\varphi}^q(\mathbf{z}) &= [\varphi_1^q(\mathbf{z}), \ldots, \varphi_{M_q}^q(\mathbf{z})]^T \end{aligned} \tag{4.12}$$

contain the approximation basis forms. The flow variables are understood as time derivatives of the distributed conserved quantities with negative sign, see

(2.37). Thus, they are approximated in the same spatial bases,

$$\begin{aligned} p_h(\mathbf{z}) &= \sum_{k=1}^{N_p} p_k \psi_k^p(\mathbf{z}) = \langle \mathbf{p} | \boldsymbol{\psi}^p(\mathbf{z}) \rangle \in \Psi_h^p, \\ q_h(\mathbf{z}) &= \sum_{l=1}^{N_q} q_l \psi_l^q(\mathbf{z}) = \langle \mathbf{q} | \boldsymbol{\psi}^q(\mathbf{z}) \rangle \in \Psi_h^q, \end{aligned} \tag{4.13}$$

and

$$\mathbf{p} = [p_1, \ldots, p_{N_p}]^T, \qquad \mathbf{q} = [q_1, \ldots, q_{N_q}]^T \tag{4.14}$$

denote the vectors of *discrete* or *integral conserved quantities*.

The mixed Galerkin approximation (4.10) of (4.5) is *exact* for flow and effort forms in the approximation spaces (4.6), (4.7) (in these subspaces, the residual error vanishes), if the following *compatibility conditions* hold:

$$\begin{aligned} \operatorname{span}\{\psi_1^p, \ldots, \psi_{N_p}^p\} &= \operatorname{span}\{\mathrm{d}\varphi_1^q, \ldots, \mathrm{d}\varphi_{M_q}^q\}, \\ \operatorname{span}\{\psi_1^q, \ldots, \psi_{N_q}^q\} &= \operatorname{span}\{\mathrm{d}\varphi_1^p, \ldots, \mathrm{d}\varphi_{M_p}^p\}. \end{aligned} \tag{4.15}$$

In contrast to [71] (Assumptions 3 and 7), this *compatibility of forms*[8] is understood *in the weak sense*. This means, more precisely – consider the original weak formulation (4.1) and the definition of the weak exterior derivative – that for all test forms with compact support inside Ω, i. e. $v^p \in H_0^1\Lambda^{n-p}(\Omega)$, $v^q \in H_0^1\Lambda^{n-q}(\Omega)$, there exist constants $a_k^p, a_l^q, b_i^p, b_j^q$ such that

$$\begin{aligned} \sum_{k=1}^{N_p} a_k^p \langle v^p | \psi_k^p \rangle_\Omega + \sum_{j=1}^{M_q} b_j^q \langle v^p | \mathrm{d}\varphi_j^q \rangle_\Omega = 0, \\ \sum_{l=1}^{N_q} a_l^q \langle v^q | \psi_l^q \rangle_\Omega + \sum_{i=1}^{M_p} b_i^p \langle v^q | \mathrm{d}\varphi_i^p \rangle_\Omega = 0. \end{aligned} \tag{4.16}$$

4.2.3 Discretized Structure Equations

We approximate the weak formulation (4.5) of the Stokes-Dirac structure by replacing the flow and effort forms with their finite-dimensional approximations (4.8), (4.9). By choosing the test forms from the effort bases,

$$v_h^p = \langle \mathbf{v}^p | \boldsymbol{\varphi}^p \rangle, \quad v_h^q = \langle \mathbf{v}^q | \boldsymbol{\varphi}^q \rangle, \qquad \mathbf{v}^p \in \mathbb{R}^{M_p},\ \mathbf{v}^q \in \mathbb{R}^{M_q}, \tag{4.17}$$

the finite-dimensional inner products in the approximation will retain the interpretation in terms of *power*. We obtain (the exterior derivative applies element-

[8] In other words, this is the de Rham property of the sequence of approximation subspaces.

wise to a vector of differential forms)

$$\begin{aligned}
\Big\langle \langle \mathbf{v}^p|\boldsymbol{\varphi}^p\rangle \;\Big|\; \langle \mathbf{f}^p|\boldsymbol{\psi}^p\rangle \Big\rangle_\Omega \;-\; & (-1)^{r+q}\Big\langle \langle \mathbf{v}^p|\mathrm{d}\boldsymbol{\varphi}^p\rangle \;\Big|\; \langle \mathbf{e}^q|\boldsymbol{\varphi}^q\rangle \Big\rangle_\Omega \\
& + (-1)^{r+q}\sum_{\mu=1}^{n_\Gamma}\Big\langle \langle \mathbf{v}^p|\boldsymbol{\varphi}^p\rangle \;\Big|\; \langle \mathbf{e}^q|\boldsymbol{\varphi}^q\rangle \Big\rangle_{\Gamma_\mu} \qquad (4.18a)\\
& + (-1)^{r+q}\sum_{\nu=1}^{\hat{n}_\Gamma}\Big\langle \langle \mathbf{v}^p|\boldsymbol{\varphi}^p\rangle \;\Big|\; \langle \mathbf{e}^q|\boldsymbol{\varphi}^q\rangle \Big\rangle_{\hat{\Gamma}_\nu} \;=\; 0,\\
\Big\langle \langle \mathbf{v}^q|\boldsymbol{\varphi}^q\rangle \;\Big|\; \langle \mathbf{f}^q|\boldsymbol{\psi}^q\rangle \Big\rangle_\Omega \;-\; & (-1)^{p}\Big\langle \langle \mathbf{v}^q|\mathrm{d}\boldsymbol{\varphi}^q\rangle \;\Big|\; \langle \mathbf{e}^p|\boldsymbol{\varphi}^p\rangle \Big\rangle_\Omega \\
& + (-1)^{p}\sum_{\mu=1}^{n_\Gamma}\Big\langle \langle \mathbf{v}^q|\boldsymbol{\varphi}^q\rangle \;\Big|\; \langle \mathbf{e}^p|\boldsymbol{\varphi}^p\rangle \Big\rangle_{\Gamma_\mu} \qquad (4.18b)\\
& + (-1)^{p}\sum_{\nu=1}^{\hat{n}_\Gamma}\Big\langle \langle \mathbf{v}^q|\boldsymbol{\varphi}^q\rangle \;\Big|\; \langle \mathbf{e}^p|\boldsymbol{\varphi}^p\rangle \Big\rangle_{\hat{\Gamma}_\nu} \;=\; 0.
\end{aligned}$$

Evaluating the integrals over the exterior products of basis forms, the system of equations can be written

$$\begin{aligned}
\Big\langle \mathbf{v}^p \Big| \mathbf{M}_p\mathbf{f}^p \Big\rangle \;+\; \Big\langle \mathbf{v}^p \Big| \Big(\mathbf{K}_p + \sum_{\mu=1}^{n_\Gamma}\mathbf{L}_p^\mu + \sum_{\nu=1}^{\hat{n}_\Gamma}\hat{\mathbf{L}}_p^\nu\Big)\mathbf{e}^q \Big\rangle = 0,\\
\Big\langle \mathbf{v}^q \Big| \mathbf{M}_q\mathbf{f}^q \Big\rangle \;+\; \Big\langle \mathbf{v}^q \Big| \Big(\mathbf{K}_q + \sum_{\mu=1}^{n_\Gamma}\mathbf{L}_q^\mu + \sum_{\nu=1}^{\hat{n}_\Gamma}\hat{\mathbf{L}}_q^\nu\Big)\mathbf{e}^p \Big\rangle = 0,
\end{aligned} \qquad (4.19)$$

with the coefficient matrices $\mathbf{M}_p \in \mathbb{R}^{M_p\times N_p}$, $\mathbf{M}_q \in \mathbb{R}^{M_q\times N_q}$, $\mathbf{K}_p, \mathbf{L}_p^\mu, \hat{\mathbf{L}}_p^\nu \in \mathbb{R}^{M_p\times M_q}$, $\mathbf{K}_q, \mathbf{L}_q^\mu, \hat{\mathbf{L}}_q^\nu \in \mathbb{R}^{M_q\times M_p}$, $\mu = 1,\dots,n_\Gamma$, $\nu = 1,\dots,\hat{n}_\Gamma$, composed of the elements

$$\begin{aligned}
[\mathbf{M}_p]_{ik} &= \langle \varphi_i^p|\psi_k^p\rangle_\Omega, & [\mathbf{M}_q]_{jl} &= \langle \varphi_j^q|\psi_l^q\rangle_\Omega,\\
[\mathbf{K}_p]_{ij} &= -(-1)^{r+q}\langle \mathrm{d}\varphi_i^p|\varphi_j^q\rangle_\Omega, & [\mathbf{K}_q]_{ji} &= -(-1)^{p}\langle \mathrm{d}\varphi_j^q|\varphi_i^p\rangle_\Omega,\\
[\mathbf{L}_p^\mu]_{ij} &= (-1)^{r+q}\langle \varphi_i^p|\varphi_j^q\rangle_{\Gamma_\mu}, & [\mathbf{L}_q^\mu]_{ji} &= (-1)^{p}\langle \varphi_j^q|\varphi_i^p\rangle_{\Gamma_\mu},\\
[\hat{\mathbf{L}}_p^\nu]_{ij} &= (-1)^{r+q}\langle \varphi_i^p|\varphi_j^q\rangle_{\hat{\Gamma}_\nu}, & [\hat{\mathbf{L}}_q^\nu]_{ji} &= (-1)^{p}\langle \varphi_j^q|\varphi_i^p\rangle_{\hat{\Gamma}_\nu}.
\end{aligned} \qquad (4.20)$$

The equations of (4.19) have to hold for arbitrary $\mathbf{v}^p \in \mathbb{R}^{M_p}$, $\mathbf{v}^q \in \mathbb{R}^{M_q}$, which yields the system of equations for the discrete flow and effort vectors

$$\begin{aligned}
\mathbf{M}_p\mathbf{f}^p + (\mathbf{K}_p + \mathbf{L}_p)\mathbf{e}^q = \mathbf{0},\\
\mathbf{M}_q\mathbf{f}^q + (\mathbf{K}_q + \mathbf{L}_q)\mathbf{e}^p = \mathbf{0}.
\end{aligned} \qquad (4.21)$$

By skew-symmetry of the wedge product, see Eq. (A.1), it is straightforward to show that

$$[\mathbf{L}_p^\mu]_{ij} = [\mathbf{L}_q^\mu]_{ji}, \qquad [\hat{\mathbf{L}}_p^\nu]_{ij} = [\hat{\mathbf{L}}_q^\mu]_{ji}, \qquad (4.22)$$

i. e. $\mathbf{L}_p^\mu = (\mathbf{L}_q^\mu)^T$ and $\hat{\mathbf{L}}_p^\nu = (\hat{\mathbf{L}}_q^\nu)^T$. By defining

$$\mathbf{L}_p = \sum_{\mu=1}^{n_\Gamma} \mathbf{L}_p^\mu + \sum_{\nu=1}^{\hat{n}_\Gamma} \hat{\mathbf{L}}_p^\nu, \qquad \mathbf{L}_q = \sum_{\mu=1}^{n_\Gamma} \mathbf{L}_q^\mu + \sum_{\nu=1}^{\hat{n}_\Gamma} \hat{\mathbf{L}}_q^\nu, \tag{4.23}$$

we can show the following.

Lemma 4.1. The matrices $\mathbf{K}_p, \mathbf{K}_q$ and $\mathbf{L}_p, \mathbf{L}_q$ are related via $[\mathbf{K}_p + \mathbf{L}_p]_{ij} + [\mathbf{K}_q + \mathbf{L}_q]_{ji} = [\mathbf{L}_p]_{ij} = [\mathbf{L}_q]_{ji}$, i. e.

$$\mathbf{L}_p \;=\; (\mathbf{K}_p + \mathbf{L}_p) + (\mathbf{K}_q + \mathbf{L}_q)^T \;=\; \mathbf{L}_q^T. \tag{4.24}$$

Proof. By the definition (4.23) and the corresponding parts of (4.20), the elements of $\mathbf{L}_p, \mathbf{L}_q$ are duality products over the effort basis forms on the complete boundary $\partial\Omega$. Thus, we have that

$$\begin{aligned}[\mathbf{K}_p + \mathbf{L}_p]_{ij} + [\mathbf{K}_q + \mathbf{L}_q]_{ji} = &-(-1)^{r+q}\langle \mathrm{d}\varphi_i^p | \varphi_j^q \rangle_\Omega + (-1)^{r+q}\langle \varphi_i^p | \varphi_j^q \rangle_{\partial\Omega} \\ &- (-1)^p \langle \mathrm{d}\varphi_j^q | \varphi_i^p \rangle_\Omega + (-1)^p \langle \varphi_j^q | \varphi_i^p \rangle_{\partial\Omega}.\end{aligned} \tag{4.25}$$

Using skew-symmetry of the wedge product (A.1) and the integration-by-parts formula for differential forms (A.7), the right hand side can be rewritten as

$$(-1)^p \langle \varphi_j^q | \varphi_i^p \rangle_{\partial\Omega} = [\mathbf{L}_q]_{ji} = [\mathbf{L}_p]_{ij}, \tag{4.26}$$

which proves the claim. □

Definition 4.2. The quadratic forms over the discrete effort vectors with the corresponding matrices $\mathbf{L}_p, \mathbf{L}_p^\mu, \hat{\mathbf{L}}_p^\nu$ and $\mathbf{L}_q, \mathbf{L}_q^\mu, \hat{\mathbf{L}}_q^\nu$ describe the approximate power transmitted over the boundary $\partial\Omega$ or its parts. We refer to these matrices as *boundary power matrices*.

The boundary power matrices $\mathbf{L}_p = \mathbf{L}_q^T$, will have reduced rank. The reason is that basis forms for interior effort degrees of freedom will be, in general, zero on the boundary. This is true e. g. for finite elements, see Section 4.5, and also for the 1D geometric pseudo-spectral collocation method [138].

4.2.4 Discrete Boundary Port Variables

To define the *pairs of discrete boundary port variables* that will be assigned either the role of boundary controls or the role of outputs on the boundary subsets, we characterize mappings on the spaces of discrete efforts variables.

Definition 4.3. The vectors of *discrete boundary port variables*[9] $\mathbf{e}^{b,\mu}, \mathbf{f}_0^{b,\mu} \in \mathbb{R}^{M_b^\mu}$ and $\hat{\mathbf{e}}^{b,\nu}, \hat{\mathbf{f}}_0^{b,\nu} \in \mathbb{R}^{\hat{M}_b^\nu}$, associated with the boundary subdomains $\Gamma_\mu \subset \partial\Omega$, $\mu = 1, \ldots, n_\Gamma$, $\hat{\Gamma}_\nu \subset \partial\Omega$, $\nu = 1, \ldots, \hat{n}_\Gamma$, satisfy

$$\langle \mathbf{e}^q | \mathbf{L}_q^\mu \mathbf{e}^p \rangle =: \langle \mathbf{e}^{b,\mu} | \mathbf{f}_0^{b,\mu} \rangle, \qquad \langle \mathbf{e}^p | \hat{\mathbf{L}}_p^\nu \mathbf{e}^q \rangle =: \langle \hat{\mathbf{e}}^{b,\nu} | \hat{\mathbf{f}}_0^{b,\nu} \rangle, \tag{4.27}$$

i. e. their duality products (which are standard Euclidean scalar products on the finite-dimensional bond space) match the discrete expression of the power flow over Γ_μ and $\hat{\Gamma}_\nu$, respectively.

We decompose the boundary power matrices for each boundary subdomain in matrix products

$$\mathbf{L}_q^\mu = (\mathbf{T}_q^\mu)^T \mathbf{S}_{p,0}^\mu, \qquad \hat{\mathbf{L}}_p^\nu = (\hat{\mathbf{T}}_p^\nu)^T \hat{\mathbf{S}}_{q,0}^\nu. \tag{4.28}$$

The *boundary trace matrices*[10] $\mathbf{T}_q^\mu \in \mathbb{R}^{M_b^\mu \times M_q}$, $\hat{\mathbf{T}}_p^\nu \in \mathbb{R}^{\hat{M}_b^\nu \times M_p}$ define the effort degrees of freedom

$$\mathbf{e}^{b,\mu} = \mathbf{T}_q^\mu \mathbf{e}^q, \qquad \hat{\mathbf{e}}^{b,\nu} = \hat{\mathbf{T}}_p^\nu \mathbf{e}^p \tag{4.29}$$

that lie on the boundary and are assigned the roles of *input variables*. The elements of $\mathbf{T}_q^\mu$ and $\hat{\mathbf{T}}_p^\nu$ are typically only zero and ± 1, depending on the orientation of the boundary. We call $\mathbf{S}_{p,0}^\mu \in \mathbb{R}^{M_b^\mu \times M_p}$, $\hat{\mathbf{S}}_{q,0}^\nu \in \mathbb{R}^{\hat{M}_b^\nu \times M_q}$ the *collocated boundary output* matrices. They define the boundary flow variables

$$\mathbf{f}_0^{b,\mu} = \mathbf{S}_{p,0}^\mu \mathbf{e}^p, \qquad \hat{\mathbf{f}}_0^{b,\nu} = \hat{\mathbf{S}}_{q,0}^\nu \mathbf{e}^q, \tag{4.30}$$

which, together with the discrete efforts (4.29), satisfy *exactly* the discrete power balance (4.27) on the different portions of the boundary[11]. Because of

$$\begin{aligned}
\langle e_\mu^\Gamma | f_\mu^\Gamma \rangle_{\Gamma_\mu} &= (-1)^p \langle e^q | e^p \rangle_{\Gamma_\mu} \\
&\approx (-1)^p \Big\langle \langle \mathbf{e}^q | \boldsymbol{\varphi}^q \rangle \Big| \langle \mathbf{e}^p | \boldsymbol{\varphi}^p \rangle \Big\rangle_{\Gamma_\mu} = \langle \mathbf{e}^q | \mathbf{L}_q^\mu \mathbf{e}^p \rangle = \langle \mathbf{e}^{b,\mu} | \mathbf{f}_0^{b,\mu} \rangle, \\
\langle \hat{e}_\nu^\Gamma | \hat{f}_\nu^\Gamma \rangle_{\hat{\Gamma}_\nu} &= (-1)^p \langle e^p | e^q \rangle_{\hat{\Gamma}_\nu} \\
&\approx (-1)^p \Big\langle \langle \mathbf{e}^p | \boldsymbol{\varphi}^p \rangle \Big| \langle \mathbf{e}^q | \boldsymbol{\varphi}^q \rangle \Big\rangle_{\hat{\Gamma}_\nu} = \langle \mathbf{e}^p | \hat{\mathbf{L}}_p^\nu \mathbf{e}^q \rangle = \langle \hat{\mathbf{e}}^{b,\nu} | \hat{\mathbf{f}}_0^{b,\nu} \rangle,
\end{aligned} \tag{4.31}$$

the definition of discrete boundary port variables is consistent with the distributed definition (4.4). Summation over the individual boundary power matrices according to (4.23) yields a matrix equation that expresses the boundary power balance,

$$\mathbf{L}_p \;=\; \mathbf{S}_{p,0}^T \mathbf{T}_q + \hat{\mathbf{T}}_p^T \hat{\mathbf{S}}_{q,0} \;=\; \mathbf{L}_q^T, \tag{4.32}$$

[9] *Discrete* boundary variables have index b, in contrast to index ∂ for the original distributed quantities on $\partial\Omega$.

[10] This denomination refers to the trace theorem for the extension of a H^m function to the boundary.

[11] The subscript 0 indicates that these discrete output variables will be re-defined when we derive a PH state space model based on a (non-degenerate) Dirac structure.

where

$$\mathbf{T}_q = \begin{bmatrix} \mathbf{T}_q^1 \\ \vdots \\ \mathbf{T}_q^{n_\Gamma} \end{bmatrix}, \quad \mathbf{S}_{p,0} = \begin{bmatrix} \mathbf{S}_{p,0}^1 \\ \vdots \\ \mathbf{S}_{p,0}^{n_\Gamma} \end{bmatrix}, \qquad \hat{\mathbf{T}}_p = \begin{bmatrix} \hat{\mathbf{T}}_p^1 \\ \vdots \\ \hat{\mathbf{T}}_p^{\hat{n}_\Gamma} \end{bmatrix}, \quad \hat{\mathbf{S}}_{q,0} = \begin{bmatrix} \hat{\mathbf{S}}_{q,0}^1 \\ \vdots \\ \hat{\mathbf{S}}_{q,0}^{\hat{n}_\Gamma} \end{bmatrix}. \tag{4.33}$$

The overall vectors of discrete boundary port variables comprise the contributions of each boundary subset with corresponding causality[12],

$$\mathbf{e}^b = \mathbf{T}_q \mathbf{e}^q, \quad \mathbf{f}_0^b = \mathbf{S}_{p,0} \mathbf{e}^p, \qquad \hat{\mathbf{e}}^b = \hat{\mathbf{T}}_p \mathbf{e}^p, \quad \hat{\mathbf{f}}_0^b = \hat{\mathbf{S}}_{q,0} \mathbf{e}^q, \tag{4.34}$$

with

$$\mathbf{e}^b = \begin{bmatrix} \mathbf{e}^{b,1} \\ \vdots \\ \mathbf{e}^{b,n_\Gamma} \end{bmatrix}, \quad \mathbf{f}_0^b = \begin{bmatrix} \mathbf{f}_0^{b,1} \\ \vdots \\ \mathbf{f}_0^{b,n_\Gamma} \end{bmatrix}, \qquad \hat{\mathbf{e}}^b = \begin{bmatrix} \mathbf{e}^{b,1} \\ \vdots \\ \mathbf{e}^{b,\hat{n}_\Gamma} \end{bmatrix}, \quad \hat{\mathbf{f}}_0^b = \begin{bmatrix} \mathbf{f}_0^{b,1} \\ \vdots \\ \mathbf{f}_0^{b,\hat{n}_\Gamma} \end{bmatrix}. \tag{4.35}$$

4.2.5 Power Balance on the Discrete Bond Space

The vectors of discrete flows and efforts $\mathbf{f}^{p/q}$, $\mathbf{e}^{p/q}$ that satisfy (4.21), together with the discrete boundary ports of different causality, define a subset of the bond space

$$\mathcal{F} \times \mathcal{E} \;=\; (\mathbb{R}^{N_p} \times \mathbb{R}^{N_q} \times \mathbb{R}^{M_b} \times \mathbb{R}^{\hat{M}_b}) \;\times\; (\mathbb{R}^{M_p} \times \mathbb{R}^{M_q} \times \mathbb{R}^{M_b} \times \mathbb{R}^{\hat{M}_b}), \tag{4.36}$$

with $M_b = \sum_{\mu=1}^{n_\Gamma} M_b^\mu$, $\hat{M}_b = \sum_{\nu=1}^{\hat{n}_\Gamma} \hat{M}_b^\nu$. On this subspace, a discrete power balance holds that approximates the continuous one (4.4).

Proposition 4.2. The subspace

$$D = \{(\mathbf{f}^p, \mathbf{f}^q, \mathbf{f}_0^b, \hat{\mathbf{f}}_0^b, \mathbf{e}^p, \mathbf{e}^q, \mathbf{e}^b, \hat{\mathbf{e}}^b) \in \mathcal{F} \times \mathcal{E} \mid (4.21) \text{ holds}\}, \tag{4.37}$$

with the boundary port variables defined by (4.29) and (4.30), satisfies the isotropy condition $D \subset D^\perp$ with respect to the bilinear form $\langle\langle \cdot, \cdot \rangle\rangle_M$ that results from symmetrization of the duality product

$$\langle \cdot | \cdot \rangle_M := \langle \mathbf{e}^p | \mathbf{M}_p \mathbf{f}^p \rangle + \langle \mathbf{e}^q | \mathbf{M}_q \mathbf{f}^q \rangle + \langle \mathbf{e}^b | \mathbf{f}_0^b \rangle + \langle \hat{\mathbf{e}}^b | \hat{\mathbf{f}}_0^b \rangle. \tag{4.38}$$

Proof. The proposition generalizes Proposition 18 in [138] and follows the same lines. We write out the symmetrized bilinear form, replacing (4.21) (short notation: $(\mathbf{K}+\mathbf{L})_{p/q} := \mathbf{K}_{p/q} + \mathbf{L}_{p/q}$):

$$\begin{aligned} &\langle \mathbf{e}_1^p | \mathbf{M}_p \mathbf{f}_2^p \rangle + \langle \mathbf{e}_1^q | \mathbf{M}_q \mathbf{f}_2^q \rangle + \langle \mathbf{e}_1^b | (\mathbf{f}_0^b)_2 \rangle + \langle \hat{\mathbf{e}}_1^b | (\hat{\mathbf{f}}_0^b)_2 \rangle \\ &\qquad + \langle \mathbf{e}_2^p | \mathbf{M}_p \mathbf{f}_1^p \rangle + \langle \mathbf{e}_2^q | \mathbf{M}_q \mathbf{f}_1^q \rangle + \langle \mathbf{e}_2^b | (\mathbf{f}_0^b)_1 \rangle + \langle \hat{\mathbf{e}}_2^b | (\hat{\mathbf{f}}_0^b)_1 \rangle \\ &= -\langle \mathbf{e}_1^p | (\mathbf{K}+\mathbf{L})_p \mathbf{e}_2^q \rangle - \langle \mathbf{e}_2^q | (\mathbf{K}+\mathbf{L})_p \mathbf{e}_1^p \rangle - \langle \mathbf{e}_1^q | (\mathbf{K}+\mathbf{L})_p \mathbf{e}_2^p \rangle - \langle \mathbf{e}_2^p | (\mathbf{K}+\mathbf{L})_p \mathbf{e}_1^q \rangle \\ &\qquad + \langle \mathbf{T}_q \mathbf{e}_1^q | \mathbf{S}_{p,0} \mathbf{e}_2^p \rangle + \langle \hat{\mathbf{T}}_p \mathbf{e}_1^p | \hat{\mathbf{S}}_{q,0} \mathbf{e}_2^q \rangle + \langle \mathbf{T}_q \mathbf{e}_2^q | \mathbf{S}_{p,0} \mathbf{e}_1^p \rangle + \langle \hat{\mathbf{T}}_p \mathbf{e}_2^p | \hat{\mathbf{S}}_{q,0} \mathbf{e}_1^q \rangle. \end{aligned} \tag{4.39}$$

[12] The causality of a pair of port variables changes if the role of in- and output is permuted.

Exploiting the matrix equalities (4.24) and (4.32), we obtain

$$-\langle \mathbf{e}_1^p | \mathbf{L}_p \mathbf{e}_2^q \rangle - \langle \mathbf{e}_1^q | \mathbf{L}_q \mathbf{e}_2^p \rangle + \langle \mathbf{e}_1^q | \mathbf{L}_q \mathbf{e}_2^p \rangle + \langle \mathbf{e}_1^p | \mathbf{L}_p \mathbf{e}_2^q \rangle = 0, \tag{4.40}$$

which proves isotropy of D with respect to $\langle\langle \cdot, \cdot \rangle\rangle_M$. □

The discrete power continuity equation, which represents the counterpart of (4.4) *in the approximation subspaces*, finally reads

$$\langle \mathbf{e}^p | \mathbf{M}_p \mathbf{f}^p \rangle + \langle \mathbf{e}^q | \mathbf{M}_q \mathbf{f}^q \rangle + \langle \mathbf{e}^b | \mathbf{f}_0^b \rangle + \langle \hat{\mathbf{e}}^b | \hat{\mathbf{f}}_0^b \rangle = 0. \tag{4.41}$$

The subspace (4.37) is, however, *not* a Dirac structure, as the duality product $\langle \cdot | \cdot \rangle_M$ defined in (4.38) is *degenerate* in general. Its value can be zero for nonzero discrete flows and/or efforts that lie in the kernel of $\mathbf{M}_p$, $\mathbf{M}_q$, or their transposes. This motivates the introduction of *power-preserving mappings* on the discrete bond space in Section 4.3.

Remark 4.3. The problem of a degenerate duality product does not appear in the approach according to [56], which is based on a *metric-dependent* Dirac structure. The parameters in the power-preserving maps introduced below represent, however, *degrees of freedom to tune* the resulting numerical methods.

4.2.6 Discrete Conservation Laws

Assume the matrices in the second terms of (4.21) can be factorized as

$$\begin{aligned} \mathbf{K}_p + \mathbf{L}_p &= -(-1)^r \mathbf{M}_p \mathbf{d}_p \\ \mathbf{K}_q + \mathbf{L}_q &= -\mathbf{M}_q \mathbf{d}_q. \end{aligned} \tag{4.42}$$

Then the set of linear equations that relates discrete flow and effort degrees of freedom has the form

$$\begin{bmatrix} \mathbf{f}^p \\ \mathbf{f}^q \end{bmatrix} = \begin{bmatrix} \mathbf{0} & (-1)^r \mathbf{d}_p \\ \mathbf{d}_q & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}^p \\ \mathbf{e}^q \end{bmatrix}. \tag{4.43}$$

This is a *direct discrete representation* of the two conservation laws with $\mathbf{d}_p \in \mathbb{R}^{N_p \times M_q}$ and $\mathbf{d}_q \in \mathbb{R}^{N_q \times M_p}$ *discrete derivative matrices* that replace the exterior derivative in the distributed parameter setting. For a mixed FE approximation based on Whitney forms of lowest polynomial degree, see e. g. [21], the representation (4.43) is obtained by integrating only over the respective discrete, oriented geometric objects (volumes, faces or edges) on the discretization mesh instead of the whole domain Ω. The matrices $\mathbf{d}_p$ and $\mathbf{d}_q$ are then the transposed incidence matrices[13], which relate the geometric objects on the mesh.

Remark 4.4. The direct discrete representation (4.43) is immediately obtained if, instead of (4.17), the test forms in the finite-dimensional approximation are chosen as

$$v_h^p = \langle \mathbf{v}^p | * \boldsymbol{\psi}^p \rangle, \quad v_h^q = \langle \mathbf{v}^q | * \boldsymbol{\psi}^q \rangle, \qquad \mathbf{v}^p \in \mathbb{R}^{N_p},\ \mathbf{v}^q \in \mathbb{R}^{N_q}. \tag{4.44}$$

[13] In order to avoid confusion with the actuated system *boundary*, we use, as in [174] or [198], the term *incidence* matrix instead of *boundary* matrix.

4.3 Power-Preserving Mappings and Dirac Structure

The discrete power balance (4.41) contains the duality pairings $\langle \mathbf{e}^p | \mathbf{M}_p \mathbf{f}^p \rangle$ and $\langle \mathbf{e}^q | \mathbf{M}_q \mathbf{f}^q \rangle$, which are *degenerate* in general, i. e. the matrices $\mathbf{M}^p$ and $\mathbf{M}^q$ may be non-quadratic and have reduced rank, see Table 4.4 for the 2D example considered in Section 4.7. To obtain a finite-dimensional *Dirac structure* with *non-degenerate* power pairings, as a basis for the PH approximation model in *state space form*, we introduce *power-preserving* mappings of the discrete flow and effort vectors onto finite-dimensional spaces of appropriate, identical dimension. We motivate these mappings by the following example.

Example 4.1. Consider the discrete power balance, a simplified representation of (4.41), $\langle \mathbf{e} | \mathbf{M}\mathbf{f} \rangle + \langle \mathbf{e}^b | \mathbf{f}_0^b \rangle = 0$ with the *degenerate* bilinear form $\langle \mathbf{e} | \mathbf{M}\mathbf{f} \rangle$. Let $\mathbf{e} \in \mathbb{R}^{n_e}$, $\mathbf{f} \in \mathbb{R}^{n_f}$, $n_f \neq n_e$ and the matrix $\mathbf{M}$ be of reduced rank $r_M < \min(n_e, n_f)$. Now choose r_M vectors $\mathbf{e}_i$ and $\mathbf{f}_i$, $i = 1, \ldots, r_M$, such that the image spaces of $\mathbf{M}$ and $\mathbf{M}^T$ are spanned by

$$\begin{aligned} \operatorname{span}\{\mathbf{M}\mathbf{f}_1, \ldots, \mathbf{M}\mathbf{f}_{r_M}\} =: \operatorname{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_{r_M}\} &= \operatorname{im}(\mathbf{M}), \\ \operatorname{span}\{\mathbf{M}^T\mathbf{e}_1, \ldots, \mathbf{M}^T\mathbf{e}_{r_M}\} =: \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{r_M}\} &= \operatorname{im}(\mathbf{M}^T). \end{aligned} \tag{4.45}$$

Suppose that the matrix $\mathbf{M}$ can be decomposed as

$$\mathbf{M} = \mathbf{P}_e^T \mathbf{P}_f \qquad \text{with} \qquad \mathbf{P}_e = \begin{bmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_{r_M}^T \end{bmatrix}, \quad \mathbf{P}_f = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_{r_M}^T \end{bmatrix}, \tag{4.46}$$

then the degenerate bilinear form can be replaced by the *non-degenerate* duality product $\langle \tilde{\mathbf{e}} | \tilde{\mathbf{f}} \rangle$ with $\tilde{\mathbf{e}} = \mathbf{P}_e \mathbf{e}$, $\tilde{\mathbf{f}} = \mathbf{P}_f \mathbf{f}$, and the discrete power balance becomes $\langle \tilde{\mathbf{e}} | \tilde{\mathbf{f}} \rangle + \langle \mathbf{e}^b | \mathbf{f}_0^b \rangle = 0$. By the definition of the rows of $\mathbf{P}_e$ and $\mathbf{P}_f$, i. e. $\mathbf{w}_i^T = \mathbf{f}_i^T \mathbf{M}^T$ and $\mathbf{v}_i^T = \mathbf{e}_i^T \mathbf{M}$, it is easy to see that $\mathbf{P}_e \mathbf{e} = \mathbf{0}$ for $\mathbf{e} \in \ker(\mathbf{M}^T)$ and $\mathbf{P}_f \mathbf{f} = \mathbf{0}$ for $\mathbf{f} \in \ker(\mathbf{M})$. This means that $\mathbf{P}_e$ and $\mathbf{P}_f$ describe mappings from the quotient spaces $\mathbb{R}^{n_e} / \ker(\mathbf{M}^T)$ and $\mathbb{R}^{n_f} / \ker(\mathbf{M})$ to $\mathbb{R}^{r_M}$, which map the equivalence classes[14]

$$\begin{aligned} [\mathbf{e}] &= \{\mathbf{e}' \in \mathbb{R}^{n_e} \mid \exists \mathbf{e}'' \in \ker(\mathbf{M}^T),\ \mathbf{e}' = \mathbf{e} + \mathbf{e}''\} \quad \text{and} \\ [\mathbf{f}] &= \{\mathbf{f}' \in \mathbb{R}^{n_f} \mid \exists \mathbf{f}'' \in \ker(\mathbf{M}),\ \mathbf{f}' = \mathbf{f} + \mathbf{f}''\} \end{aligned} \tag{4.47}$$

onto an embedding of $\mathbb{R}^{n_e} \times \mathbb{R}^{n_f}$, endowed with coordinates $(\tilde{\mathbf{e}}, \tilde{\mathbf{f}})$. We call $\tilde{\mathbf{e}}, \tilde{\mathbf{f}} \in \mathbb{R}^{\tilde{N}}$ *minimal* discrete power variables with $\tilde{N} = r_M$ in the considered case.

If no factorization (4.46) exists – this is the case if the dimension of the minimal bond variables is lower than the rank of $\mathbf{M}$, $\tilde{N} < r_M$ – the "internal" power term $\langle \mathbf{e} | \mathbf{M}\mathbf{f} \rangle$ can not be matched with $\langle \tilde{\mathbf{e}} | \tilde{\mathbf{f}} \rangle$. Preservation of the total

[14] The maps from $\mathbb{R}^{n_e}$ and $\mathbb{R}^{n_f}$ to the quotient spaces are *projections*.

discrete power balance will in such a case be achieved by an appropriate re-definition of the output $\mathbf{f}_0^b \to \mathbf{f}^b$ such that $\langle \tilde{\mathbf{e}} | \tilde{\mathbf{f}} \rangle + \langle \mathbf{e}^b | \mathbf{f}^b \rangle = 0$ holds, see the following subsection. For an illustration, consider Example 4.5: The original output vector $\hat{\mathbf{f}}_0^b$ does *not* contain the rotational components contained in $\hat{\mathbf{f}}^b$ as depicted in Fig. 4.13.

4.3.1 Minimal Discrete Bond Variables

We use the argumentation sketched above to construct a *Dirac structure* on a *minimal* discrete bond space. To replace $\langle \mathbf{e}^p | \mathbf{M}_p \mathbf{f}^p \rangle$ and $\langle \mathbf{e}^q | \mathbf{M}_q \mathbf{f}^q \rangle$ in (4.41) by *non-degenerate* duality pairings, we determine *power-preserving mappings*

$$\tilde{\mathbf{e}}^p = \mathbf{P}_{ep} \mathbf{e}^p, \quad \tilde{\mathbf{e}}^q = \mathbf{P}_{eq} \mathbf{e}^q \qquad \text{and} \qquad \tilde{\mathbf{f}}^p = \mathbf{P}_{fp} \mathbf{f}^p, \quad \tilde{\mathbf{f}}^q = \mathbf{P}_{fq} \mathbf{f}^q, \tag{4.48}$$

such that

$$\begin{aligned} \tilde{N}_p &:= \dim \tilde{\mathbf{e}}^p = \dim \tilde{\mathbf{f}}^p \leq \operatorname{rank}(\mathbf{M}_p) \qquad \text{and} \\ \tilde{N}_q &:= \dim \tilde{\mathbf{e}}^q = \dim \tilde{\mathbf{f}}^q \leq \operatorname{rank}(\mathbf{M}_q). \end{aligned} \tag{4.49}$$

We refer to the vectors $\tilde{\mathbf{f}}^p, \tilde{\mathbf{e}}^p \in \mathbb{R}^{\tilde{N}_p}$, $\tilde{\mathbf{f}}^q, \tilde{\mathbf{e}}^q \in \mathbb{R}^{\tilde{N}_q}$ as *minimal* discrete flows and efforts, as they can be interpreted as coordinates of an embedding in the original discrete bond space.

Example 4.2. In the 1D case, $p = q = 1$, using Whitney finite elements or the pseudo-spectral method [138], we have, $N = N_p = N_q$ and $M = M_p = M_q$ with $M = N + 1$. Fixing $\tilde{\mathbf{f}}^p = \mathbf{f}^p$, $\tilde{\mathbf{f}}^q = \mathbf{f}^q$, minimal discrete efforts can be defined as $\tilde{\mathbf{e}}^p = \mathbf{M}_p^T \mathbf{e}^p$ and $\tilde{\mathbf{e}}^q = \mathbf{M}_q^T \mathbf{e}^q$.

The following definition summarizes the core property of power-preserving mappings.

Definition 4.4. The discrete flow and effort mappings (4.48) are called *power-preserving* if they satisfy a discrete power balance

$$\langle \tilde{\mathbf{e}}^p | \tilde{\mathbf{f}}^p \rangle + \langle \tilde{\mathbf{e}}^q | \tilde{\mathbf{f}}^q \rangle + \langle \mathbf{e}^b | \mathbf{f}^b \rangle + \langle \hat{\mathbf{e}}^b | \hat{\mathbf{f}}^b \rangle = 0 \tag{4.50}$$

with the given boundary inputs $\mathbf{e}^b$, $\hat{\mathbf{e}}^b$ according to (4.29) and possibly modified boundary outputs

$$\mathbf{f}^b = \mathbf{S}_p^\mu \mathbf{e}^p, \qquad \hat{\mathbf{f}}^b = \hat{\mathbf{S}}_q^\nu \mathbf{e}^q. \tag{4.51}$$

Remark 4.5. If the mappings satisfy $\mathbf{P}_{ep}^T \mathbf{P}_{fp} = \mathbf{M}_p$ and $\mathbf{P}_{eq}^T \mathbf{P}_{fq} = \mathbf{M}_q$, the "interior" part of the power balance (4.41) is exactly represented by the minimal flows $\tilde{\mathbf{f}}$ and efforts $\tilde{\mathbf{e}}$, and (4.50) holds with the original, collocated outputs $\mathbf{f}^b = \mathbf{f}_0^b$, $\hat{\mathbf{f}}^b = \hat{\mathbf{f}}_0^b$. If, however, $\tilde{N}_q < \operatorname{rank}(\mathbf{M}_q)$ and/or $\tilde{N}_p < \operatorname{rank}(\mathbf{M}_p)$, a part of the power, originally described by $\langle \mathbf{e}^p | \mathbf{M}_p \mathbf{f}^p \rangle + \langle \mathbf{e}^q | \mathbf{M}_q \mathbf{f}^q \rangle$, must be "swapped" to the boundary terms of (4.50) via the re-definition of the outputs. This way, the power-balance is maintained globally, and *conservativeness* of the finite-dimensional approximation is *guaranteed*.

To characterize the power-preserving mappings and modified output maps that guarantee power continuity (4.50), we substitute in this equation the definitions of the effort and flow maps, the in- and outputs, and substitute $\mathbf{f}^p$, $\mathbf{f}^q$ according to the discrete representation (4.43) of the conservation laws. The new power variables are now expressed in terms of the original discrete efforts,

$$\underbrace{\begin{bmatrix}\tilde{\mathbf{f}}^p\\ \hat{\mathbf{f}}^b\\ \tilde{\mathbf{f}}^q\\ \mathbf{f}^b\end{bmatrix}}_{\bar{\mathbf{f}}} = \underbrace{\begin{bmatrix}\mathbf{0} & (-1)^r\mathbf{P}_{fp}\mathbf{d}_p\\ \mathbf{0} & \hat{\mathbf{S}}_q\\ \mathbf{P}_{fq}\mathbf{d}_q & \mathbf{0}\\ \mathbf{S}_p & \mathbf{0}\end{bmatrix}}_{\mathbf{E}^T}\underbrace{\begin{bmatrix}\mathbf{e}^p\\ \mathbf{e}^q\end{bmatrix}}_{\mathbf{e}}, \qquad \underbrace{\begin{bmatrix}\tilde{\mathbf{e}}^p\\ \hat{\mathbf{e}}^b\\ \tilde{\mathbf{e}}^q\\ \mathbf{e}^b\end{bmatrix}}_{\bar{\mathbf{e}}} = \underbrace{\begin{bmatrix}\mathbf{P}_{ep} & \mathbf{0}\\ \hat{\mathbf{T}}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{eq}\\ \mathbf{0} & \mathbf{T}_q\end{bmatrix}}_{\mathbf{F}^T}\underbrace{\begin{bmatrix}\mathbf{e}^p\\ \mathbf{e}^q\end{bmatrix}}_{\mathbf{e}}. \tag{4.52}$$

Equation (4.50) must hold for arbitrary $\mathbf{e}^p$, $\mathbf{e}^q$, and by substituting the components of $\bar{\mathbf{f}}$ and $\bar{\mathbf{e}}$ in (4.50), we obtain the following matrix condition.

Proposition 4.3 (Power-preserving mappings)**.** The effort, flow and output maps are *power-preserving*, if they satisfy the matrix equation

$$(-1)^r\mathbf{d}_p^T\mathbf{P}_{fp}^T\mathbf{P}_{ep} + \mathbf{P}_{eq}^T\mathbf{P}_{fq}\mathbf{d}_q + \mathbf{T}_q^T\mathbf{S}_p + \hat{\mathbf{S}}_q^T\hat{\mathbf{T}}_p = \mathbf{0}. \tag{4.53}$$

The power-preserving maps are not unique. Different parametrizations of the matrices yield different finite-dimensional Dirac structures that approximate the original Stokes-Dirac structure as defined in Theorem 2.4. Together with a consistent approximation of the constitutive equations, we obtain PH approximate models with different numerical properties. A favorable parametrization will depend on the nature of the system (e. g. if dynamics and closure equations make the system hyperbolic or parabolic), the distribution and type of boundary inputs, and the application case. In any case, the power-preserving maps generate a *minimal* space of power variables on which an approximate Dirac structure is defined.

In Sections 4.6 and 4.7, we will illustrate the construction of the power-preserving maps on the example of Whitney approximation forms in 1D and on a rectangular simplicial mesh in 2D. The degrees of freedom in the mappings will allow for a trade-off between centered schemes and upwinding in the discretized PH models.

Remark 4.6. Equation (4.53) relates the "discrete differentiation matrices" $\mathbf{d}_p$, $\mathbf{d}_q$ and the "discrete trace matrices" $\mathbf{T}_q$, $\hat{\mathbf{T}}_p$, paired with $\mathbf{S}_p$, $\hat{\mathbf{S}}_q$. This is an apparent reference to Stokes' theorem A.1, which is instrumental in deriving this discrete representation of power continuity (see also Eq. (43) in [138]).

4.3.2 Dirac Structure

The power-preserving maps that satisfy (4.53) define a *Dirac structure*. We verify that (4.52) is an *image representation* of this Dirac structure on the

minimal discrete bond space. If the effort maps are invertible, an unconstrained input-output representation exists.

Proposition 4.4 (Image representation)**.** Consider the discrete flow and effort vectors $\bar{\mathbf{f}}$ and $\bar{\mathbf{e}}$ as indicated in (4.52). $(\bar{\mathbf{f}}, \bar{\mathbf{e}})$ is an element of the bond space

$$\bar{\mathcal{F}} \times \bar{\mathcal{E}} = \mathbb{R}^{\tilde{N}_p + \hat{M}_b + \tilde{N}_q + M_b} \times \mathbb{R}^{\tilde{N}_p + \hat{M}_b + \tilde{N}_q + M_b}. \tag{4.54}$$

Let $\tilde{N}_p + \tilde{N}_q + \hat{M}_b + M_b = M_p + M_q$ and assume that the matrix condition (4.53) is satisfied. If

$$\operatorname{rank}\left(\begin{bmatrix}\mathbf{P}_{ep}\\ \hat{\mathbf{T}}_p\end{bmatrix}\right) = M_p \qquad \text{and} \qquad \operatorname{rank}\left(\begin{bmatrix}\mathbf{P}_{eq}\\ \mathbf{T}_q\end{bmatrix}\right) = M_q, \tag{4.55}$$

then the subspace defined by (4.52), i. e.

$$\bar{D} = \{(\bar{\mathbf{f}}, \bar{\mathbf{e}}) \in \bar{\mathcal{F}} \times \bar{\mathcal{E}} \mid \bar{\mathbf{f}} = \mathbf{E}^T\mathbf{e},\ \bar{\mathbf{e}} = \mathbf{F}^T\mathbf{e},\ \mathbf{e} \in \mathbb{R}^{M^p + M^q}\}, \tag{4.56}$$

is a Dirac structure.

Proof. According to the image representation of a Dirac structure, see Theorem 2.2, the dimensions of $\bar{\mathbf{f}}$ and $\bar{\mathbf{e}}$ must be less[15] or equal $\dim(\mathbf{e})$, which is ensured by $\tilde{N}_p + \tilde{N}_q + \hat{M}_b + M_b = M_p + M_q$. The condition $\operatorname{rank}([\mathbf{F}\ \ \mathbf{E}]) = M_p + M_q$ is satisfied by (4.55), from which $\operatorname{rank}(\mathbf{F}) = M_p + M_q$ follows. Moreover, the skew-symmetry condition $\mathbf{E}\mathbf{F}^T + \mathbf{F}\mathbf{E}^T = \mathbf{0}$ must hold. $\mathbf{E}\mathbf{F}^T + \mathbf{F}\mathbf{E}^T$ according to (4.52) gives

$$\begin{bmatrix}\mathbf{0} & \mathbf{X}^T\\ \mathbf{X} & \mathbf{0}\end{bmatrix}, \quad \mathbf{X} = (-1)^r \mathbf{d}_p^T \mathbf{P}_{fp}^T \mathbf{P}_{ep} + \mathbf{P}_{eq}^T \mathbf{P}_{fq} \mathbf{d}_q + \mathbf{T}_q^T \mathbf{S}_p + \hat{\mathbf{S}}_q^T \hat{\mathbf{T}}_p. \tag{4.57}$$

Because of (4.53), this is the zero matrix, which completes the proof. □

4.4 Finite-Dimensional Port-Hamiltonian Model

The fact that both matrices in (4.55) are assumed square and invertible guarantees the existence of an explicit *input-output representation* of the above-defined Dirac structure.

Corollary 4.1 (Input-output representation)**.** Under the conditions of Proposition 4.4, the Dirac structure admits an *unconstrained input-output representation*

$$\begin{bmatrix}-\tilde{\mathbf{f}}^p\\ -\tilde{\mathbf{f}}^q\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{J}_p\\ \mathbf{J}_q & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{e}}^p\\ \tilde{\mathbf{e}}^q\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{B}_p\\ \mathbf{B}_q & \mathbf{0}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{e}}^b\\ \mathbf{e}^b\end{bmatrix}, \tag{4.58a}$$

$$\begin{bmatrix}\hat{\mathbf{f}}^b\\ \mathbf{f}^b\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{C}_q\\ \mathbf{C}_p & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{e}}^p\\ \tilde{\mathbf{e}}^q\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{D}_q\\ \mathbf{D}_p & \mathbf{0}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{e}}^b\\ \mathbf{e}^b\end{bmatrix} \tag{4.58b}$$

with

$$\mathbf{J}_p = -\mathbf{J}_q^T, \quad \mathbf{C}_q = \mathbf{B}_q^T, \quad \mathbf{C}_p = \mathbf{B}_p^T, \quad \mathbf{D}_q = -\mathbf{D}_p^T. \tag{4.59}$$

[15] This is the case of a *relaxed* image representation.

Proof. The latter (skew-)symmetry conditions can be summarized as

$$\begin{bmatrix} -\mathbf{J}_p & -\mathbf{B}_p \\ \mathbf{C}_q & \mathbf{D}_q \end{bmatrix} + \begin{bmatrix} -\mathbf{J}_q & -\mathbf{B}_q \\ \mathbf{C}_p & \mathbf{D}_p \end{bmatrix}^T = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{4.60}$$

The submatrices in Eq. (4.58) are obtained from evaluation of $\bar{\mathbf{f}}$ in (4.52) and exploiting invertibility of the matrices in (4.55). We can write

$$\begin{aligned} \begin{bmatrix} -\mathbf{J}_p & -\mathbf{B}_p \\ \mathbf{C}_q & \mathbf{D}_q \end{bmatrix} &= \begin{bmatrix} (-1)^r \mathbf{P}_{fp} \mathbf{d}_p \\ \hat{\mathbf{S}}_q \end{bmatrix} \begin{bmatrix} \mathbf{P}_{eq} \\ \mathbf{T}_q \end{bmatrix}^{-1}, \\ \begin{bmatrix} -\mathbf{J}_q & -\mathbf{B}_q \\ \mathbf{C}_p & \mathbf{D}_p \end{bmatrix} &= \begin{bmatrix} \mathbf{P}_{fq} \mathbf{d}_q \\ \mathbf{S}_p \end{bmatrix} \begin{bmatrix} \mathbf{P}_{ep} \\ \hat{\mathbf{T}}_p \end{bmatrix}^{-1}. \end{aligned} \tag{4.61}$$

Substituting these relations in (4.60) and multiplying with the non-singular matrices $\begin{bmatrix} \mathbf{P}_{ep}^T & \hat{\mathbf{T}}_p^T \end{bmatrix}$ from the left and $\begin{bmatrix} \mathbf{P}_{eq} \\ \mathbf{T}_q \end{bmatrix}$ from the right yields the left hand side of (4.53). The right hand side being zero, this proves (skew-)symmetry of the matrices (4.59) of the input-output representation. □

Remark 4.7. The proposition is a generalization of Proposition 20 in [138] for the 1D case and the pseudo-spectral method. Note that the rank condition (4.55) on the effort and boundary maps, which is expressed in addition to (4.53), is sufficient (not necessary) for the subspace (4.56) to be a Dirac structure.

To build from the input-output representation of the Dirac structure a finite-dimensional PH model for the *canonical system of two conservation laws*, we replace the minimal discrete flow variables by time derivatives of *discrete states*

$$-\tilde{\mathbf{f}}^p =: \dot{\tilde{\mathbf{p}}} \in \mathbb{R}^{\tilde{N}_p}, \qquad -\tilde{\mathbf{f}}^q =: \dot{\tilde{\mathbf{q}}} \in \mathbb{R}^{\tilde{N}_q}. \tag{4.62}$$

To complete the geometric discretization, a *consistent approximation* of the constitutive equations is necessary. For the considered hyperbolic case, the minimal efforts need to be expressed as the partial derivatives of a suitable discrete Hamiltonian $\tilde{H}_d(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$,

$$\tilde{\mathbf{e}}^p = \left(\frac{\partial \tilde{H}_d}{\partial \tilde{\mathbf{p}}} \right)^T \in \mathbb{R}^{\tilde{N}_p}, \qquad \tilde{\mathbf{e}}^q = \left(\frac{\partial \tilde{H}_d}{\partial \tilde{\mathbf{q}}} \right)^T \in \mathbb{R}^{\tilde{N}_q}. \tag{4.63}$$

We present the discretization of the constitutive equations in more detail in the FE examples of the following sections.

With the combined state, input and output vectors

$$\mathbf{x} = \begin{bmatrix} \tilde{\mathbf{p}} \\ \tilde{\mathbf{q}} \end{bmatrix} \in \mathbb{R}^{\tilde{N}_p + \tilde{N}_q}, \quad \mathbf{u} = \begin{bmatrix} \hat{\mathbf{e}}^b \\ \mathbf{e}^b \end{bmatrix} \in \mathbb{R}^{M_b + \hat{M}_b}, \quad \mathbf{y} = \begin{bmatrix} \hat{\mathbf{f}}^b \\ \mathbf{f}^b \end{bmatrix} \in \mathbb{R}^{M_b + \hat{M}_b}, \tag{4.64}$$

the resulting state space model has explicit PH form ($\mathbf{J} = -\mathbf{J}^T$, $\mathbf{D} = -\mathbf{D}^T$)

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{J}\nabla H_d(\mathbf{x}) + \mathbf{B}\mathbf{u} \\ \mathbf{y} &= \mathbf{B}^T\nabla H_d(\mathbf{x}) + \mathbf{D}\mathbf{u} \end{aligned} \tag{4.65}$$

and the discrete energy satisfies the balance equation

$$\dot{H}_d = -\mathbf{y}^T\mathbf{u}, \tag{4.66}$$

which is the finite-dimensional counterpart of (2.44). The PH form allows to easily interconnect the finite-dimensional model of the system of two conservation laws with other subsystems in a power-preserving way, which is the basis for energy-based control design *by interconnection* see e. g. [122].

4.5 Whitney Finite Elements

We will show the application of the introduced structure-preserving discretization approach using *Whitney forms* [209] of lowest polynomial degree to set up the finite element approximation bases for flows and efforts (4.6) and (4.7). Whitney forms can be constructed based on the *barycentric* node weights [23]. The degrees of freedom (in 3D) are directly associated to the *nodes*, oriented *edges*, *faces* and *volumes* of the simplicial discretization mesh. The geometric discretization of Maxwell's equations as described in [21] is based on Whitney forms, and the resulting finite-dimensional models feature the (co-)incidence matrices of the underlying discretization meshes [24]. They can be considered a *direct representation* of the physical laws on the discrete balance regions of the triangulation.

Whitney forms in 1D. Consider a one-dimensional domain $\Omega = (0, L) \subset \mathbb{R}$, which is divided by equidistant nodes (step size $h = \frac{L}{N}$) $z_i = (i-1)h$, $i = 1, \ldots, N+1$, into N intervals $I_k = ((k-1)h, kh)$, $k = 1, \ldots, N$. The Whitney 0-forms (node forms) over the discretization grid are the well known "hat functions"

$$w^{n_i} = \begin{cases} \frac{1}{h}(z - z_{i-1}), & z \in I_{i-1}, \quad i > 1, \\ 1 - \frac{1}{h}(z - z_i), & z \in I_i, \qquad i < N+1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.67}$$

The Whitney 1-forms (edge forms)

$$w^{e_k} = \begin{cases} \frac{1}{h}dz, & z \in I_k, \\ 0, & \text{otherwise,} \end{cases} \tag{4.68}$$

are piecewise constant and satisfy $\int_\Omega w^{e_k} = 1$, $k = 1, \ldots, N$.

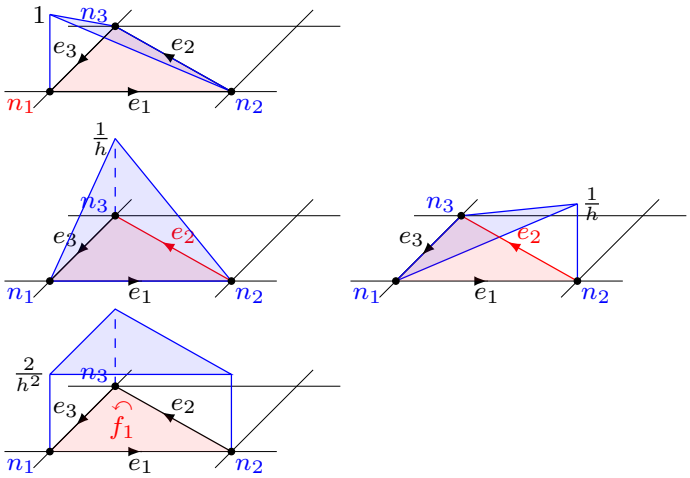

Figure 4.1: Illustrations of the Whitney forms over a 2-simplex. Top: Node form w^{n_1}. Middle: Negative x- and positive y-component of the edge form w^{e_2}. Bottom: Face form w^{f_1}.

Whitney forms over a 2D simplex. Consider the triangle $f_1 = \{(x, y) \mid x, y \geq 0,\ 0 \leq x + y \leq h\}$, with vertices $n_1 = (0, 0)$, $n_2 = (h, 0)$, $n_3 = (0, h)$, which are connected by the oriented edges e_1, e_2 and e_3 as shown in Fig. 4.1. The node, edge and face forms are constructed according to [23]:

$$w^{n_1} = 1 - \frac{x}{h} - \frac{y}{h}, \quad w^{n_2} = \frac{x}{h}, \quad w^{n_3} = \frac{y}{h}, \tag{4.69}$$

$$w^{e_1} = \frac{h-y}{h^2}dx + \frac{x}{h^2}dy, \quad w^{e_2} = -\frac{y}{h^2}dx + \frac{x}{h^2}dy, \quad w^{e_3} = -\frac{y}{h^2}dx + \frac{x-h}{h^2}dy, \tag{4.70}$$

$$w^{f_1} = \frac{2}{h^2}dx \wedge dy. \tag{4.71}$$

The 0-, 1- and 2-forms verify $w^{n_i}(n_j) = \delta_{ij}$, $\int_{e_i} w^{e_j} = \delta_{ij}$ (δ_{ij} the Kronecker-Delta) and $\int_{f_1} w^{f_1} = 1$.

4.6 One-Dimensional Examples

In this section, we apply the structure-preserving discretization method introduced above to the benchmark 1D examples of the linear wave and the linear heat equation on the interval $\Omega = (0, 1) \subset \mathbb{R}$. We use the Whitney node and edge forms to approximate efforts and flows. We discuss the quality of

the resulting numerical schemes in terms of the eigenvalue approximation and illustrate it by the numerical solutions of initial value problems. The power-preserving mappings in the 1D case feature a single tuning parameter, which allows to adapt the approximation method to the nature of the considered system. While *upwinding* will be convenient in the hyperbolic case, a centered scheme is favorable for a parabolic system.

4.6.1 Discretization of the Structure Equations

Both the wave and the heat equation share the same structure equation, see (2.59a) and (2.88a). For $n = 1$, we have

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & \mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix} \tag{4.72}$$

with $f^p, f^q \in L^2\Lambda^1(\Omega)$ and $e^p, e^q \in H^1\Lambda^0(\Omega)$. We consider the (input) boundary efforts

$$\hat{e}^\partial = \operatorname{tr} e^p|_{z=1}\,, \qquad e^\partial = \operatorname{tr} e^q|_{z=0}\,. \tag{4.73}$$

The collocated and power-conjugated boundary flows are

$$\hat{f}^\partial = -\operatorname{tr} e^q|_{z=1}\,, \qquad f^\partial = \operatorname{tr} e^p|_{z=0}\,. \tag{4.74}$$

It is straightforward to verify that with this choice, the structural power balance

$$\langle e^p | f^p \rangle_\Omega + \langle e^q | f^q \rangle_\Omega + \langle e^\partial | f^\partial \rangle_{\partial\Omega} + \langle \hat{e}^\partial | \hat{f}^\partial \rangle_{\partial\Omega} = 0 \tag{4.75}$$

holds. Using the weak form of the structure equations (4.72) and the finite-dimensional effort and flow spaces $\Phi^p = \Phi^q$ and $\Psi^p = \Psi^q$, which are spanned by Whitney node and edge forms according to (4.67) and (4.68), we obtain a linear system of equations in the form

$$\begin{aligned} \mathbf{M}_p \mathbf{f}^p &= (\mathbf{K}_p + \mathbf{L}_p)\mathbf{e}^q, \\ \mathbf{M}_q \mathbf{f}^q &= (\mathbf{K}_q + \mathbf{L}_q)\mathbf{e}^p. \end{aligned} \tag{4.76}$$

The degrees of freedom in $\mathbf{f}^p, \mathbf{f}^q \in \mathbb{R}^N$ are approximations of the integrals of $f^p, f^q \in L^2\Lambda^1(\Omega)$ over the intervals I_k, while the elements of $\mathbf{e}^p, \mathbf{e}^q \in \mathbb{R}^{N+1}$ approximate the node values of $e^p, e^q \in H^1\Lambda^0(\Omega)$. The matrices in (4.76) are, according to (4.20),

$$\mathbf{M}_p = \mathbf{M}_q = \frac{1}{2} \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \\ & & & 1 \end{bmatrix} \in \mathbb{R}^{(N+1)\times N}, \tag{4.77}$$

$$\mathbf{K}_p = \mathbf{K}_q = \frac{1}{2}\begin{bmatrix} -1 & -1 & & & \\ 1 & 0 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & -1 \\ & & & 1 & 1 \end{bmatrix} \in \mathbb{R}^{(N+1)\times(N+1)} \tag{4.78}$$

and

$$\mathbf{L}_p = \mathbf{L}_q = \begin{bmatrix} 1 & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & -1 \end{bmatrix} \in \mathbb{R}^{(N+1)\times(N+1)}. \tag{4.79}$$

The sums $(\mathbf{K}_p + \mathbf{L}_p)$ and $(\mathbf{K}_q + \mathbf{L}_q)$ can be factorized according to (4.42), such that (4.76) can be rewritten as the discrete conservation laws

$$\begin{bmatrix} \mathbf{f}^p \\ \mathbf{f}^q \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{d}_p \\ \mathbf{d}_q & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}^p \\ \mathbf{e}^q \end{bmatrix}. \tag{4.80}$$

with

$$\mathbf{d}_p = \mathbf{d}_q = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{N\times(N+1)} \tag{4.81}$$

the co-incidence matrices of the oriented discretization grid. Note that $\mathbf{d}_p$ and $\mathbf{d}_q$ can be directly obtained using the test functions indicated in (4.44).

4.6.2 Power-Preserving Mappings

The definition of boundary efforts (4.73) translates into the $1 \times (N+1)$ input trace matrices

$$\mathbf{T}_q = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}, \qquad \hat{\mathbf{T}}_p = \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix}. \tag{4.82}$$

By the $N \times (N+1)$ matrices

$$\mathbf{P}_{eq} = \begin{bmatrix} \mathbf{0}_{N\times 1} & \mathbf{I}_N \end{bmatrix}, \qquad \mathbf{P}_{ep} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_{N\times 1} \end{bmatrix}, \tag{4.83}$$

those effort degrees of freedom are identified, which play the role of co-states $\tilde{e}_i^p$, $\tilde{e}_j^q$, $i, j = 1, \dots, N$, in the discretized model. The matrices in (4.55) become permutation matrices and condition (4.53) can be written

$$\begin{bmatrix} \mathbf{d}_p^T \mathbf{P}_{fp}^T & \hat{\mathbf{S}}_q^T \end{bmatrix} + \begin{bmatrix} \mathbf{S}_p \\ \mathbf{P}_{fq}\mathbf{d}_q \end{bmatrix} = \mathbf{0}. \tag{4.84}$$

The $N \times N$ flow mapping matrices

$$\mathbf{P}_{fp} = \mathbf{P}_{fq}^T = \begin{bmatrix} 1-\alpha & & & \\ \alpha & 1-\alpha & & \\ & \ddots & \ddots & \\ & & \alpha & 1-\alpha \end{bmatrix} \tag{4.85}$$

and the $1 \times (N+1)$ power-conjugated output matrices

$$\mathbf{S}_p = \begin{bmatrix} 1-\alpha & \alpha & 0 & \dots & 0 \end{bmatrix}, \qquad \hat{\mathbf{S}}_q = \begin{bmatrix} 0 & \dots & 0 & -\alpha & \alpha-1 \end{bmatrix}, \tag{4.86}$$

which satisfy (4.84), contain a single parameter $\alpha \in \mathbb{R}$. It serves as a degree of freedom in our discretization approach. Its effect on the resulting PH state space models, in particular on the quality of the numerical approximations for the wave and the heat equation, will be discussed and illustrated in the following subsections.

4.6.3 Constitutive Equations

Both linear wave and heat equation feature constitutive equations that close the corresponding system representation. In the sequel, we consider cases with normalized (material) parameters. For the 1D wave equation, according to (2.59c) and with the Hamiltonian

$$H = \frac{1}{2}\langle p | * p \rangle_\Omega + \frac{1}{2}\langle q | * q \rangle_\Omega, \tag{4.87}$$

we have the linear relations

$$e^p = \delta_p H = * p, \qquad e^q = \delta_q H = * q. \tag{4.88}$$

A similar situation occurs for the heat equation with constant heat conductivity and heat capacity $\lambda = c_v = 1$, see (2.88c),

$$e^p = * p, \qquad e^q = - * f^q, \tag{4.89}$$

with the difference that the second equation relates effort and flow.

In both cases, we look for lumped approximations of the Hodge star operator. We determine *diagonal Hodge matrices* (see e. g. [178]) such that the above constitutive equations translate into

$$\tilde{\mathbf{e}}^p = \mathbf{Q}_p \tilde{\mathbf{p}}, \qquad \tilde{\mathbf{e}}^q = \mathbf{Q}_q \tilde{\mathbf{q}} \tag{4.90}$$

and

$$\tilde{\mathbf{e}}^p = \mathbf{Q}_p \tilde{\mathbf{p}}, \qquad \tilde{\mathbf{e}}^q = -\mathbf{Q}_q \tilde{\mathbf{f}}^q, \tag{4.91}$$

respectively. $\mathbf{Q}_p$ and $\mathbf{Q}_q$ are constructed based on the requirement of *consistency in a steady state* and using the above-defined power-preserving mappings.

We derive the diagonal Hodge matrices based on the hyperbolic case of the wave equation, and apply the results accordingly to the discretized heat equation.

To this end, we compute the variation of the approximate Hamiltonian functional H^h, i. e. (4.87) under substitution of the finite-dimensional approximations

$$p^h = \sum_{k=1}^{N} p_k \psi_k, \qquad q^h = \sum_{k=1}^{N} q_k \psi_k \tag{4.92}$$

of both conserved quantities. Herein, $\psi_k = w^{e,k} \in L^2\Lambda^1(\Omega)$ denotes the Whitney edge forms to approximate p and q, respectively. The following identities hold due to $\langle w^{e,k} | * w^{e,l} \rangle_\Omega = \frac{1}{h}\delta_{kl}$, with δ_{kl} the Kronecker delta:

$$\langle \psi_k | * p^h \rangle_\Omega = p_k \langle \psi_k | * \psi_k \rangle_\Omega = \frac{p_k}{h}, \qquad \langle \psi_k | * q^h \rangle_\Omega = q_k \langle \psi_k | * \psi_k \rangle_\Omega = \frac{q_k}{h}. \tag{4.93}$$

The variations δp_k and δq_k in the discrete degrees of freedom, $k = 1, \ldots, N$, yield (neglecting higher order terms) the first variation of the approximate energy functional

$$\delta H^h = \sum_{k=1}^{N} \underbrace{\langle \psi_k | * \psi_k \rangle_\Omega \, p_k}_{=:\delta_{p_k} H^h} \delta p_k + \sum_{k=1}^{N} \underbrace{\langle \psi_k | * \psi_k \rangle_\Omega \, q_k}_{=:\delta_{q_k} H^h} \delta q_k. \tag{4.94}$$

The expressions

$$\delta_{p_k} H^h = \frac{p_k}{h}, \qquad \delta_{q_k} H^h = \frac{q_k}{h}, \tag{4.95}$$

$k = 1, \ldots, N$, represent *consistent* approximations of the continuous constitutive equations (4.88). This becomes clear when we assume uniform distributions of the conserved quantities

$$p_s = \bar{p}\, dz, \qquad q_s = \bar{q}\, dz \tag{4.96}$$

with constants $\bar{p}, \bar{q} \in \mathbb{R}$. Substitution of (4.96) in (4.88) gives $\delta_p H = \bar{p}$ and $\delta_q H = \bar{q}$. Using Whitney edge forms, the discrete degrees of freedom to realize such a constant distribution are $\bar{p}_k = h\bar{p}$, $\bar{q}_k = h\bar{q}$, $k = 1, \ldots, N$. Replacing these values in (4.95) gives finally $\delta_{p_k} H^h = \bar{p}$ and $\delta_{q_k} H^h = \bar{q}$, which can be used to define the co-state variables

$$\tilde{e}^p_k = \delta_{p_k} H^h = \frac{p_k}{h}, \qquad \tilde{e}^q_k = \delta_{q_k} H^h = \frac{q_k}{h}, \tag{4.97}$$

$k = 1, \ldots, N$, in the interior nodes of the discretization grid.

Note that p_k and q_k are elements of the original vectors $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$ of edge degrees of freedom for the conserved quantities. The discrete constitutive equations (4.90), however, are expressed in terms of the discrete state vectors

$$\tilde{\mathbf{p}} = \mathbf{P}_{fp}\mathbf{p}, \qquad \tilde{\mathbf{q}} = \mathbf{P}_{fq}\mathbf{q}. \tag{4.98}$$

In the 1D case considered here, the maps $\mathbf{P}_{fp}$ and $\mathbf{P}_{fq}$ according to (4.85) are invertible for $\alpha \neq 1$, however, the matrices $\mathbf{P}_{fp}^{-1}$ and $\mathbf{P}_{fq}^{-1}$ are lower triangular and hence not sparse for $\alpha \neq 0$. Moreover, in the 2D case presented in the next section, the mapping matrices for flows/states will not even be square. For this reason, we reverse the mappings (4.98) only for constant, identical discrete states, which is exactly the case we consider for a consistent approximation of the constitutive equations. We write (4.98) element-wise:

$$\tilde{p}_i = \sum_{i=1}^{N} [\mathbf{P}_{fp}]_{i,k} p_k, \qquad \tilde{q}_i = \sum_{i=1}^{N} [\mathbf{P}_{fq}]_{i,k} q_k. \tag{4.99}$$

For identical and constant (indicated by the bar) degrees of freedom $\bar{p}_1 = \ldots = \bar{p}_N$ and $\bar{q}_1 = \ldots = \bar{q}_N$, we can set $p_k = \bar{p}_i$ and $q_k = \bar{q}_i$, which allows to write

$$\tilde{p}_i = \left(\sum_{i=1}^{N} [\mathbf{P}_{fp}]_{i,k} \right) \bar{p}_i, \qquad \tilde{q}_i = \left(\sum_{i=1}^{N} [\mathbf{P}_{fq}]_{i,k} \right) \bar{q}_i. \tag{4.100}$$

Inversion of the latter relations and substitution in (4.97) yields the consistently computed diagonal elements, $i = 1, \ldots, N$,

$$[\mathbf{Q}_p]_{i,i} = \frac{1}{h \sum_{k=1}^{N} [\mathbf{P}_{fp}]_{i,k}}, \qquad [\mathbf{Q}_q]_{i,i} = \frac{1}{h \sum_{k=1}^{N} [\mathbf{P}_{fq}]_{i,k}}, \tag{4.101}$$

of the Hodge matrices in (4.90). For the flow/state mappings as defined in (4.85), we finally obtain

$$\mathbf{Q}_p = \frac{1}{h} \mathrm{diag}\{ \frac{1}{1-\alpha}, 1, \ldots, 1 \}, \tag{4.102a}$$

$$\mathbf{Q}_q = \frac{1}{h} \mathrm{diag}\{ 1, \ldots, 1, \frac{1}{1-\alpha} \}. \tag{4.102b}$$

In the hyperbolic case, the so-defined lumped co-states can be derived from the quadratic Hamiltonian

$$H_d(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) = \frac{1}{2} \tilde{\mathbf{p}}^T \mathbf{Q}_p \tilde{\mathbf{p}} + \frac{1}{2} \tilde{\mathbf{q}}^T \mathbf{Q}_q \tilde{\mathbf{q}}. \tag{4.103}$$

4.6.4 Interpretation of the Mapping Parameter

When using Whitney approximation forms, the meaning of the discretization degrees of freedom $p_1, \ldots, p_N$ and $q_1, \ldots, q_N$ as accumulated conserved quantities, as well as $e_1^p, \ldots, e_{N+1}^p$ and $e_1^q \ldots e_{N+1}^p$ as nodal effort variables, allows for a clear interpretation of the mapping parameter α. To this end, we consider the discretized constitutive equations for the wave equation[16] (4.90), as well as

[16] A corresponding argumentation holds for the constitutive equations (4.91) for the heat equation.

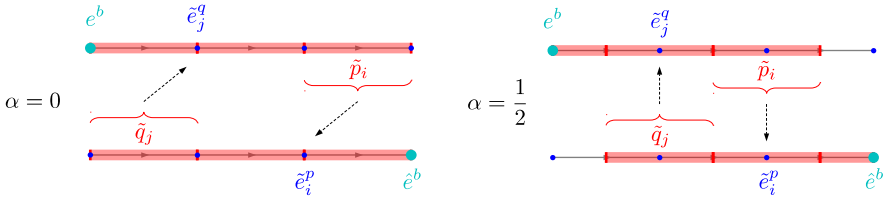

Figure 4.2: The variation of α modifies the weights of the integral conserved quantities in the definition of the state vectors $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$. This can also be interpreted as a shift of the balance intervals, which are used to compute the lumped co-states $\tilde{\mathbf{e}}^p$ und $\tilde{\mathbf{e}}^q$, which are located in the grid nodes.

the definition of lumped states $\tilde{\mathbf{p}} = \mathbf{P}_{fp}\mathbf{p}$ and $\tilde{\mathbf{q}} = \mathbf{P}_{fq}\mathbf{q}$ via the matrices in (4.85). While the elements of the vectors $\tilde{\mathbf{e}}^p = \mathbf{P}_{ep}\mathbf{e}^p$ and $\tilde{\mathbf{e}}^q = \mathbf{P}_{eq}\mathbf{e}^q$ represent approximate values of the co-state variables in the interior grid nodes (i. e. the nodes where no boundary condition is imposed), the elements of $\mathbf{p}$ and $\mathbf{q}$ stand for approximate integral conserved quantities between the nodes. The elements of the state vectors $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$ are weighted sums (factors α and $1-\alpha$) of these lumped conserved quantities. From this fact, we can conclude the following interpretations of the discretized constitutive equations depending on the value of α. Figure 4.2 illustrates the effects of the parameter choices $\alpha = 0$ (Case 1) and $\alpha = \frac{1}{2}$ (Case 3) on the definition of discrete co-states.

Case 1, $\alpha = 0$

The co-state $\tilde{e}_i^p$ ($\tilde{e}_j^q$), which is located in a grid node, is computed based on the lumped state $\tilde{p}_i = p_i$ ($\tilde{q}_j = q_j$) on the neighboring interval to the right (to the left). This *one-sided* numerical approximation of the constitutive equations gives preference to the direction from where the input information comes, i. e. where the corresponding boundary effort $\hat{e}^b$ (e^b) is imposed as an input. The same holds qualitatively if $\alpha < \frac{1}{2}$. In this sense, we can understand such a choice of α as an *upwinding* parametrization.

Case 2, $\alpha < 0$

In this case, the constitutive equations can be rewritten as

$$\tilde{e}_i^p = \frac{1}{h}\left(p_i + (-\alpha)(p_i - p_{i-1})\right), \qquad \tilde{e}_j^q = \frac{1}{h}\left(q_j + (-\alpha)(q_j - q_{j+1})\right). \quad (4.104)$$

Now, $-\alpha > 0$ weights the difference between two adjacent conserved quantities. As will be evident in the following subsection, a slightly negative value of α is favorable for the numerical solution of the hyperbolic wave equation.

Case 3, $\alpha = \frac{1}{2}$

In this case, the conserved quantities to both sides are equally weighted in order to compute the node effort. Such a *centered approximation* fits to nature of the parabolic heat equation, which is characterized by *symmetric* diffusion. Subsection 4.6.6 illustrates the superiority of the centered approximation compared to a one-sided scheme for the linear heat equation.

4.6.5 Wave Equation

In this subsection, we set up the approximate finite-dimensional state space representation for the wave equation. We analyze the quality of the discretized models in terms of the eigenvalue approximation (taken from [104]), and illustrate the difference of our approach to [71]. The numerical solution of an initial value problem (as described in [97]) supports the findings from the eigenvalue analysis. The inverse boundary value problem for feedforward control will be addressed in Chapter 6.

4.6.5.1 State Space Model

The matrices defined in Subsections 4.6.1 and 4.6.2 are substituted in (4.61) in order to obtain the input-output representation of the finite-dimensional approximation Dirac structure. The in- and outputs according to (4.73) and (4.74) are

$$u_1 = e^q(0), \quad u_2 = e^p(1), \qquad y_1 = e^p(0), \quad y_2 = -e^q(1). \tag{4.105}$$

The matrices in (4.61) are

$$\mathbf{J}_p = -\mathbf{J}_q^T = \begin{bmatrix} \alpha - 1 & & & & \\ 1 - 2\alpha & \alpha - 1 & & & \\ \alpha & 1 - 2\alpha & \alpha - 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \alpha & 1 - 2\alpha & \alpha - 1 \end{bmatrix}, \tag{4.106a}$$

$$\mathbf{B}_p^T = \mathbf{C}_p = \mathbf{S}_p = \begin{bmatrix} 1 - \alpha & \alpha & 0 & \dots & 0 \end{bmatrix}, \tag{4.106b}$$

$$\mathbf{B}_q^T = \mathbf{C}_q = \hat{\mathbf{S}}_q = \begin{bmatrix} 0 & \dots & 0 & -\alpha & \alpha - 1 \end{bmatrix}, \tag{4.106c}$$

$$\mathbf{D}_p = -\mathbf{D}_q^T = 0. \tag{4.106d}$$

Remark 4.8. Note that *no structural feedthrough* occurs in our scheme, in contrast to the structure-preserving discretization approach proposed in [71].

With the constitutive equations (4.90) and the lumped dynamics

$$\dot{\tilde{\mathbf{p}}} = -\tilde{\mathbf{f}}^p, \qquad \dot{\tilde{\mathbf{q}}} = -\tilde{\mathbf{f}}^q, \tag{4.107}$$

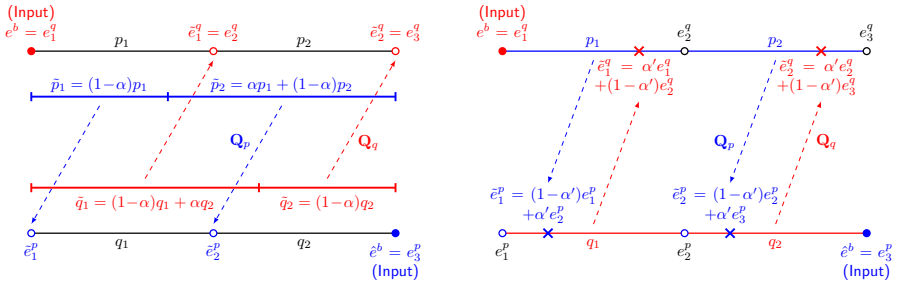

Figure 4.3: Illustration of the difference between our approach and the method according to [71]. In our approach, the discrete efforts at the *interior* nodes are computed – via appropriate discrete Hodge matrices – based on *convex sums* of the original discrete states (left sketch). Following [71], the original discrete states remain unchanged, but the *co-states* are computed as convex sums of the node efforts (right sketch). For $\alpha = 0$ and $\alpha' = 0$ both methods coincide. Then, for example, the co-state $\tilde{e}_1^q = e_2^q$ is determined based on the integral conserved quantity q_1 in both cases.

we obtain the approximate port-Hamiltonian state space model

$$\begin{bmatrix} \dot{\tilde{\mathbf{p}}} \\ \dot{\tilde{\mathbf{q}}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{J}_p \\ -\mathbf{J}_p^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_q \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{p}} \\ \tilde{\mathbf{q}} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_q \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \tag{4.108a}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_p^T & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_q^T \end{bmatrix} \begin{bmatrix} \mathbf{Q}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_q \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{p}} \\ \tilde{\mathbf{q}} \end{bmatrix} \tag{4.108b}$$

of the wave equation. Note that the model is *conservative*, i. e. no numerical dissipation is introduced by the discretization scheme, independent of the parameter α. This tuning parameter, however, can be used to influence the approximation quality of the resulting numerical models.

4.6.5.2 Eigenvalues

In this subsection, we compare the results of our method with those obtained with the approach in [71], where mapping the *efforts* at the boundary nodes of each discretization interval using a parameter[17] α' yields non-degenerate power pairings and a PH model in state space form. The *strong* compatibility conditions, which restrict the parameter value to $\alpha' = \frac{1}{2}$ for the case of lowest order Whitney forms in the original work, can be relaxed by a weak formulation of the problem. Unlike (4.102), the discrete Hodge matrices according to [71] do not depend on α': $\mathbf{Q}_p = \mathbf{Q}_q = \frac{1}{h}\mathrm{diag}\{1, \ldots, 1\}$. In contrast to our method, the state space models according to [71] feature a *direct feedthrough*[18].

[17] We use a prime to distinguish from the α in our method.

[18] The exception with zero feedthrough matrix is $\alpha' = 0$, which corresponds to $\alpha = 0$ in our approach. With these parameter values, both methods produce models that coincide

The fundamental difference between both approaches is illustrated by the sketches in Fig. 4.3 and the explanation below. For $\alpha < \frac{1}{2}$ and $\alpha' < \frac{1}{2}$, the state information from the directions in which the associated effort variables *are imposed* as boundary inputs, obtains a higher weight. This type of *upwinding* leads to a very good approximation of the eigenvalues for values close to zero of α and α'.

We consider the spectrum of the canonical differential operator of the Stokes-Dirac structure, see (4.72) under homogeneous Dirichlet boundary conditions on the efforts, i. e. (4.73) set to zero. Note that these boundary conditions correspond to Dirichlet-Neumann conditions for the PDE in second order form. The exact eigenvalues of the operator are

$$\lambda_{k,\infty} = \pm\frac{2k-1}{2}\pi i, \quad k = 1, 2, 3, \ldots, \tag{4.109}$$

see e. g. [77]. As the structure-preserving discretization is conservative, also the approximate eigenvalues have zero real parts. We display in Table 4.1 the imaginary parts for different values of the *flow* mapping parameter α. Table 4.2 shows the corresponding values for the structure-preserving discretization according to [71] with different *effort* mapping parameters α'. The relative errors for the first, 5th and 20th eigenvalue are plotted in the diagrams of Fig. 4.4.

For all displayed parametrizations around $\alpha = \alpha' = 0$, the order of the first eigenvalue approximation error is $\mathcal{O}(h)$ with $h = \frac{1}{N}$, see the top diagrams in Figs. 4.4. This is in accordance with the consistency order 1 for the *non-centered* approximation of the node efforts (see Subsection 3.4.3 for the discussion from the finite volumes point of view). We observe that for the parametrizations $\alpha = -\frac{1}{12}$ and $\alpha' = \frac{1}{12}$, the approximation quality of the higher eigenvalues is improved. The result of this favorable *upwinding* will be illustrated in the next subsection with the solution of an initial value problem[19]. Tables 4.1 and 4.2 as well as Figure 4.4 show a very similar evolution of the eigenvalues under grid refinement. Note however, that our approach, in contrast to [71], produces *no structural feedthrough*, which is *appropriate* for hyperbolic systems[20]. Moreover, the extension to 2D (and prospectively 3D) of our method is straightforward, in contrast to [71].

---

with those obtained from discrete modeling/finite volumes on regularly staggered grids [174], [95].

[19] Note that the same effect can be achieved if in the finite volume approach on regularly staggered grids [95] (which corresponds to $\alpha = 0/\alpha' = 0$), the control volumes to compute the numerical fluxes are slightly shifted.

[20] The feedthrough, together with the over-estimation of the highest eigenvalues for $\alpha' \to 0.5$, fits to the good results the method according to [71] achieves for the discretization of *parabolic* systems [10], where the instantaneous propagation of information must be approximated.

Table 4.1: Eigenvalue imaginary parts for the discretized 1D wave equation with different flow mapping parameters α and grid sizes N, compared to the exact values.

Table 4.2: A comparable approximation of the eigenvalues can be obtained using the method presented in [71] in a *weak* formulation, which permits to choose parameters other than $\alpha' = \frac{1}{2}$. Note, however, that the resulting state space models, in contrast to our approach, feature a *structural feedthrough*.

k	Exact	$\alpha = -1/12$			$\alpha = 0$			$\alpha = 1/6$		
		$N = 20$	$N = 40$	$N = 80$	$N = 20$	$N = 40$	$N = 80$	$N = 20$	$N = 40$	$N = 80$
1	1.5708	1.5263	1.5482	1.5594	1.5321	1.5513	1.5610	1.5440	1.5576	1.5642
2	4.7124	4.5798	4.6449	4.6783	4.5873	4.6516	4.6825	4.6074	4.6663	4.6910
3	7.8540	7.6352	7.7422	7.7974	7.6156	7.7449	7.8021	7.5975	7.7562	7.8131
4	10.996	10.692	10.840	10.917	10.599	10.827	10.919	10.468	10.815	10.927
5	14.137	13.746	13.939	14.036	13.521	13.892	14.031	13.176	13.830	14.030
10	29.845	28.593	29.428	29.641	26.613	28.814	29.490	23.192	27.852	29.269
20	61.261	46.492	59.316	60.828	39.883	54.899	59.422	26.831	47.252	57.198
40	124.09	–	93.244	120.76	–	79.940	111.47	–	53.664	95.338
80	249.76	–	–	186.62	–	–	159.97	–	–	107.33

k	Exact	$\alpha' = 1/12$			$\alpha' = 0$			$\alpha' = -1/6$		
		$N = 20$	$N = 40$	$N = 80$	$N = 20$	$N = 40$	$N = 80$	$N = 20$	$N = 40$	$N = 80$
1	1.5708	1.5387	1.5546	1.5627	1.5321	1.5513	1.5610	1.5189	1.5447	1.5577
2	4.7124	4.6152	4.6636	4.6879	4.5873	4.6516	4.6825	4.5283	4.6266	4.6712
3	7.8540	7.6888	7.7719	7.8130	7.6156	7.7449	7.8021	7.4544	7.6858	7.7789
4	10.996	10.757	10.879	10.938	10.599	10.827	10.919	10.250	10.708	10.877
5	14.137	13.816	13.985	14.062	13.521	13.892	14.031	12.875	13.679	13.961
10	29.845	28.700	29.459	29.675	26.613	28.814	29.490	22.886	27.377	29.052
20	61.261	47.800	59.416	60.773	39.883	54.899	59.422	29.950	46.903	56.384
40	124.09	–	95.897	120.87	–	79.940	111.47	–	59.974	94.912
80	249.76	–	–	191.95	–	–	159.97	–	–	119.99

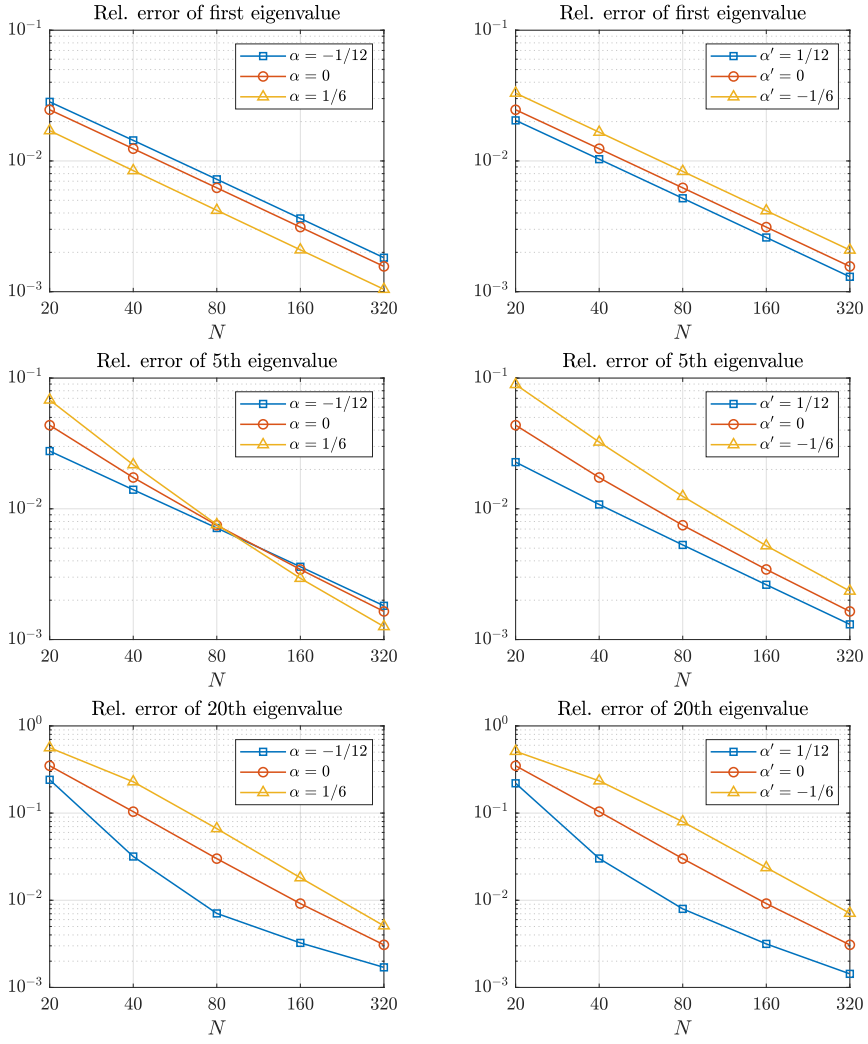

Figure 4.4: Magnitude of the approximation error over different grid sizes for the first, 5th and 20th eigenvalue of the canonical system operator. Left column: Error under the presented approach – upwinding improves the approximation of higher eigenvalues. Right column: Numerical approximation with the method according to [71].

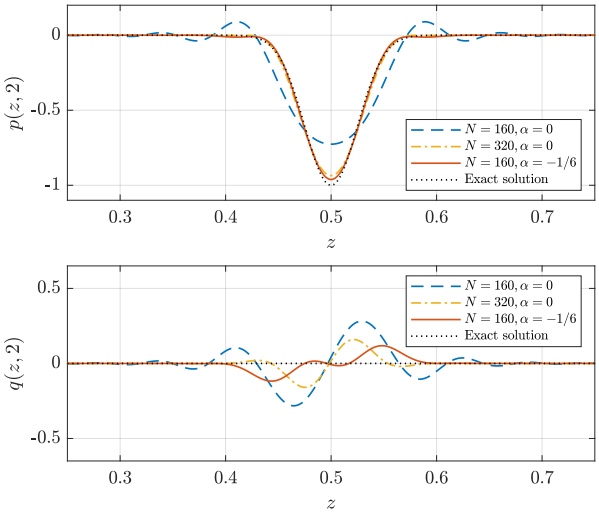

Figure 4.5: Approximations of the solution $p(z,2) = -p_0(z)$, $q(z,2) = q_0(z) = 0$ after twofold reflection on the boundaries.

4.6.5.3 Initial Value Problem

The simulation results in this subsection support the finding that slightly negative values of α are suitable for the numerical approximation of the wave equation. With $\alpha = -\frac{1}{6}$, half the number of discretization intervals is necessary to produce a comparable error as in the case of $\alpha = 0$. We consider homogeneous boundary conditions $e^{\partial} = \hat{e}^{\partial} = 0$ and an initial distribution of the state differential forms $p(t) = p(z,t)\,dz$, $q(t) = q(z,t)\,dz$ on $\Omega = (0,1)$ with

$$p(z,0) = p_0(z) = e^{-\frac{(z-0,5)^2}{0,025^2}}, \qquad q(z,0) = q_0(z) = 0. \tag{4.110}$$

With the change of variables $\eta = \frac{1}{2}(p-q)$ und $\xi = \frac{1}{2}(p+q)$, the solution on an unbounded domain consists of a left travelling and a right travelling wave

$$\eta(z,t) = \eta_0(z+t), \qquad \xi(z,t) = \xi_0(z-t), \tag{4.111}$$

which each transport the initial conditions

$$\eta_0(z) = \xi_0(z) = \frac{1}{2}p_0(z). \tag{4.112}$$

On the domain $\Omega = (0,1)$, reflections take place at the boundaries (with a change of sign at $z = 1$), such that

$$p(z,2) = -p_0(z), \qquad q(z,2) = 0 \tag{4.113}$$

is the exact solution at time $t = 2$, depicted by the dotted line in Fig. 4.5. Besides, Fig. 4.5 contains the numerical solutions with the one-sided approximations by $\alpha = 0$ with $N = 160$ (1, dashed) and $N = 320$ (2, dash-dotted), as

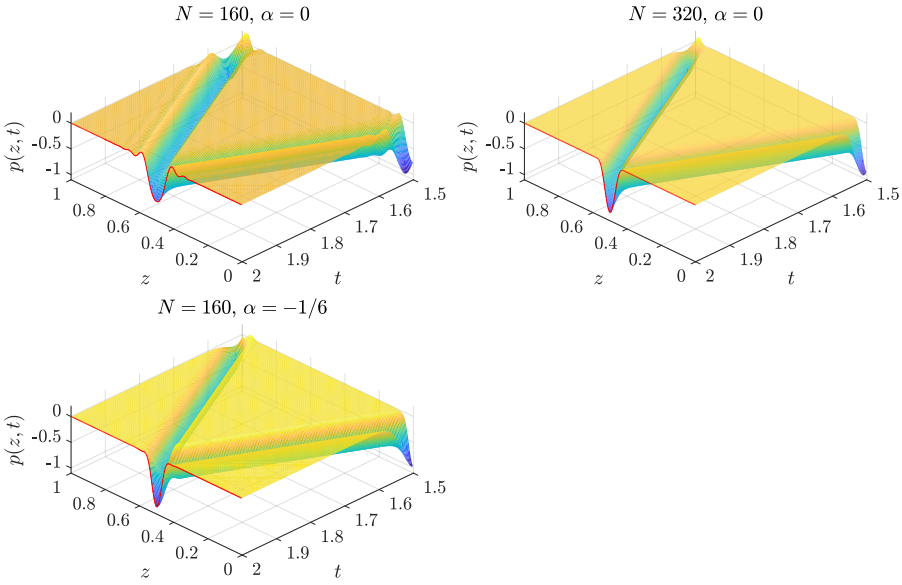

Figure 4.6: Illustration of the travelling waves in $p(z,t)$ for $N = 160$ (left), $N = 320$ (right) and $\alpha = 0$ (first row), $\alpha = -\frac{1}{6}$ (second row).

well as the numerical solution for $N = 160$ and $\alpha = -1/6$ (3, solid). The latter parametrization generates on the same grid ($N = 160$) far less dispersion than (1). It is comparable with solution (2), which is based on a finer mesh with twice the number of discretization intervals. Figure 4.6 depicts the numerical solution $p(z,t)$ over $z \in [0,1]$ and $t \in [1.5,2]$ for the different values of α and grid sizes.

4.6.6 Heat Equation

We consider the 1D heat equation, which can, see Subsection 2.3.3, be represented by the same structure equation (4.72) as the 1D wave equation. For the following study, we consider the unit interval $\Omega \in (0,1)$ and constant heat conductivity and heat capacity $\lambda = c_v = 1$. Figure 4.7 shows a sketch of the considered heat conductor with boundary conditions

$$u = \hat{e}^{\partial} = e^{p}(1), \qquad 0 = e^{\partial} = e^{q}(0). \tag{4.114}$$

A temperature is imposed at $z = 1$ as input u, while the homogeneous boundary condition at $z = 0$ represents thermal insulation (zero heat flow). This time, unlike the wave equation example, we consider the *non-collocated* output

$$y = e^{p}(0), \tag{4.115}$$

which is known to be a *flat output* for the heat equation [109]. The preservation of flatness of this output and feedforward control design based on the discretized

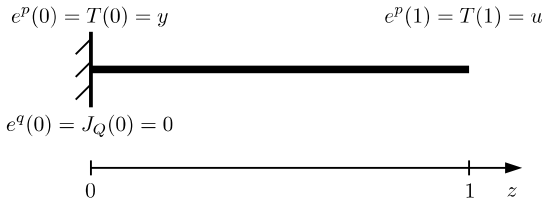

Figure 4.7: Heat conductor with Dirichlet and Neumann boundary conditions $T(1) = u$, $J_Q(0) = 0$ and flat output $T(0) = y$.

models are topics in Chapter 6. Similar to the previous subsection, we now analyze the properties of the resulting finite-dimensional state space model, which is obtained from structure-preserving discretization.

4.6.6.1 State Space Model

The state space model of the discretized heat equation with boundary conditions (4.114) is based on the same matrices as derived in Subsections 4.6.1, 4.6.2 and 4.6.3. Under the given boundary conditions, (4.58a) becomes

$$\begin{bmatrix} -\tilde{\mathbf{f}}^p \\ -\tilde{\mathbf{f}}^q \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{J}_p \\ \mathbf{J}_q & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{e}}^p \\ \tilde{\mathbf{e}}^q \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_q \end{bmatrix} u. \tag{4.116}$$

Substituting dynamics

$$\dot{\tilde{\mathbf{p}}} = -\tilde{\mathbf{f}}^p \tag{4.117}$$

and the constitutive equations

$$\tilde{\mathbf{e}}^p = \mathbf{Q}_p \tilde{\mathbf{p}}, \qquad \tilde{\mathbf{e}}^q = -\mathbf{Q}_q \tilde{\mathbf{f}}^q \tag{4.118}$$

according to (4.91), we obtain, now with the N-dimensional state vector $\mathbf{x} := \tilde{\mathbf{p}}$,

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \tag{4.119}$$

where

$$\mathbf{A} = \underbrace{-\mathbf{J}_p \mathbf{Q}_q \mathbf{J}_p^T}_{=-\mathbf{R}} \mathbf{Q}_p \in \mathbb{R}^{N \times N} \qquad \text{and} \qquad \mathbf{b} = \mathbf{J}_p \mathbf{Q}_q \mathbf{B}_q \in \mathbb{R}^N. \tag{4.120}$$

The positive definiteness of the symmetric dissipation matrix $\mathbf{R} = \mathbf{R}^T$ reflects the diffusive character of the heat equation. The discrete version of the output (4.115),

$$y = \mathbf{c}^T \mathbf{Q}_p \mathbf{x}, \tag{4.121}$$

with

$$\mathbf{c}^T = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{1 \times N} \tag{4.122}$$

completes the finite-dimensional state space model.

The state matrix $\mathbf{A}$ is symmetric and pentadiagonal,

$$\mathbf{A} = \mathrm{diag}(\mathbf{a}_0) + \mathrm{diag}_1(\mathbf{a}_1) + \mathrm{diag}_{-1}(\mathbf{a}_1) + \mathrm{diag}_2(\mathbf{a}_2) + \mathrm{diag}_{-2}(\mathbf{a}_2), \quad (4.123)$$

with the vectors of main diagonal elements $\mathbf{a}_0 \in \mathbb{R}^N$ and the elements on the first two upper and lower off-diagonals $\mathbf{a}_1 \in \mathbb{R}^{N-1}$ and $\mathbf{a}_2 \in \mathbb{R}^{N-2}$,

$$\mathbf{a}_0 = N^2 \begin{bmatrix} -1+\alpha \\ -2+6\alpha-5\alpha^2 \\ -2+6\alpha-6\alpha^2 \\ \vdots \\ -2+6\alpha-6\alpha^2 \\ -2+5\alpha-5\alpha^2 \end{bmatrix}, \quad \mathbf{a}_1 = N^2 \begin{bmatrix} 1-3\alpha+2\alpha^2 \\ 1-4\alpha+4\alpha^2 \\ \vdots \\ 1-4\alpha+4\alpha^2 \end{bmatrix}, \quad \mathbf{a}_2 = N^2 \begin{bmatrix} \alpha-\alpha^2 \\ \vdots \\ \alpha-\alpha^2 \end{bmatrix}. \quad (4.124)$$

The input vector $\mathbf{b} \in \mathbb{R}^N$ reads

$$\mathbf{b} = N \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha-\alpha^2 \\ 1-2\alpha+2\alpha^2 \end{bmatrix}. \quad (4.125)$$

$\mathbf{a}_1$ and $\mathbf{a}_2$ become zero vectors for the cases $\alpha = \frac{1}{2}$ and $\alpha = 0$, respectively. These two cases will be analyzed further below.

4.6.6.2 Eigenvalues and Zeros of the PDE Model

We compare the properties of the discretized model to those of the infinite-dimensional one. In particular, we are interested in the eigenvalues and zeros. The former give information about the approximation quality of the dynamics, the (non-)existence of the latter allows to assess whether the input $u = e^p(1)$ and the state $\mathbf{x}$ can be parametrized by the flat output $y = e^p(0)$ and its time derivatives, as in the infinite-dimensional case.

To analyze the eigenvalues of the 1D heat equation on $\Omega = (0, 1) \subset \mathbb{R}$ under homogeneous Neumann-Dirichlet boundary conditions, we write it as a second order PDE in the spatial variable z:

$$\partial_t x(z,t) = \partial_z^2 x(z,t), \qquad \partial_z x(0,t) = 0, \quad x(1,t) = u(t). \quad (4.126)$$

It is easily verified, that an initial condition in the form

$$x(z,0) = \sum_{k=1}^{\infty} c_k \cos(\sqrt{-\lambda_{k,\infty}}\, z) \quad (4.127)$$

with the negative real *eigenvalues*

$$\lambda_{k,\infty} = -\left(\frac{2k-1}{2}\pi\right)^2, \qquad k = 1, 2, \ldots \quad (4.128)$$

decays according to

$$x(z,t) = \sum_{k=1}^{\infty} c_k e^{\lambda_k t} \cos(\sqrt{-\lambda_{k,\infty}} z). \tag{4.129}$$

The Laplace transformation of the heat equation (see Appendix B.2) allows to establish the transfer function between the input $\hat{u}(s) = \hat{x}(1,s)$ and the output $\hat{y}(s) = \hat{x}(0,s)$ as

$$\hat{y}(s) = \frac{1}{\cosh(\sqrt{s})} \hat{u}(s). \tag{4.130}$$

The poles of this transfer function are exactly the eigenvalues $\lambda_{k,\infty}$ as indicated above, no zeros occur.

4.6.6.3 Approximate Eigenvalues and Zeros

Based on the structure of the matrices $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{c}^T$, we first analyze the eigenvalues and possible zeros of the state space model (4.119), (4.121) for the two cases $\alpha = 0$ and $\alpha = \frac{1}{2}$. In these two cases, the state matrix $\mathbf{A}$ becomes tridiagonal, which allows for an analytical computation of the eigenvalues. Moreover, we will observe that (after pole-zero cancellation in the second case) the transfer function does not feature (transmission) zeros, which corresponds to the infinite-dimensional system representation, see Appendix B.2.

Case 1, $\alpha = 0$

The matrices $\mathbf{A}$ and $\mathbf{b}$ have the structure

$$\mathbf{A} = N^2 \mathbf{X} = N^2 \begin{bmatrix} -1 & 1 & & \\ 1 & -2 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{bmatrix}, \qquad \mathbf{b} = N \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{4.131}$$

The roots λ_k, $k = 1, \ldots, N$, of the characteristic polynomial $p(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A})$ are the roots λ'_k of $p'(\lambda') = \det(\lambda' \mathbf{I} - \mathbf{X})$, multiplied with N^2. The tridiagonal form of $\mathbf{X}$ allows to construct the characteristic polynomial by recursion of the determinants of south-eastern submatrices. This recursion resembles the one for the Chebyshev polynomials of the first kind. With the changes of variables $\lambda' + 2 = 2\mu$ and $\mu = \cos(\xi)$, one finds the roots of p',

$$\xi_k = \frac{2k-1}{2N+1}\pi, \qquad k = 1, \ldots, N, \tag{4.132}$$

and finally the eigenvalues of $\mathbf{A}$

$$\lambda_k = 2N^2 \left(\cos(\frac{2k-1}{2N+1}\pi) - 1 \right), \qquad k = 1, \ldots, N. \tag{4.133}$$

The invariant zeros are the roots of $\det(\mathbf{P}(\eta))$ with

$$\mathbf{P}(\eta) = \begin{bmatrix} \mathbf{A} - \eta\mathbf{I} & \mathbf{b} \\ \mathbf{c}^T & 0 \end{bmatrix} \tag{4.134}$$

the Rosenbrock matrix. A short calculation shows that

$$\begin{aligned} \det(\mathbf{P}(\eta)) &= (-1)^N N^4 \det([\mathbf{X} - \eta\mathbf{I}]_{1:N-1,2:N}) \\ &= (-1)^N N^4, \end{aligned} \tag{4.135}$$

where $[\mathbf{X} - \eta\mathbf{I}]_{1:N-1,2:N}$ denotes the $(N-1) \times (N-1)$ north-eastern submatrix of $\mathbf{X} - \eta\mathbf{I}$, which is lower triangular. The discretized models $(\mathbf{A}, \mathbf{b}, \mathbf{c}^T)$ with parameter $\alpha = 0$ have *no* invariant zeros. Thus, the output has full relative degree N. It is a flat output, as the input and the states can be parametrized in terms of $y(t)$ and its time derivatives, which allows to compute a feedforward control $u(t)$ as in the infinite-dimensional case, see Chapter 6.

Case 2, $\alpha = \frac{1}{2}$

The $\mathbf{A}$ matrix has a chessboard pattern. We assume an *even* number N of discretization intervals. The non-zero elements of $\mathbf{A}$ can be collected in the matrix $\mathbf{A}_1 := [\mathbf{A}]_{1:2:N-1,1:2:N-1}$ of odd rows and columns and the matrix $\mathbf{A}_2 := [\mathbf{A}]_{2:2:N,2:2:N}$ of even rows and columns:

$$\mathbf{A}_1 = \frac{N^2}{4}\begin{bmatrix} -2 & 1 & & & \\ 2 & -2 & 1 & & \\ & 1 & -2 & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & -2 \end{bmatrix}, \qquad \mathbf{A}_2 = \frac{N^2}{4}\begin{bmatrix} -1 & 1 & & & \\ 1 & -2 & \ddots & & \\ & \ddots & \ddots & 1 & \\ & & 1 & -2 & 1 \\ & & & 1 & -3 \end{bmatrix}. \tag{4.136}$$

Both matrices are *similar*, they are related via the equality $\mathbf{A}_1\mathbf{S} = \mathbf{S}\mathbf{A}_2$ with the upper triangular matrix

$$\mathbf{S} = \begin{bmatrix} 1 & -1 & 1 & \cdots & -(-1)^N \\ & 2 & -2 & \cdots & 2(-1)^N \\ & & \ddots & \ddots & \\ & & & 2 & -2 \\ & & & & 2 \end{bmatrix}, \tag{4.137}$$

whose elements have alternating sign. Consequently, $\mathbf{A}_1$ and $\mathbf{A}_2$ have identical eigenvalues. The analysis of the characteristic polynomial of $\mathbf{A}_2$, with the help of the recurrence relation for Chebyshev polynomials, results in the eigenvalues

$$\lambda_k = \frac{N^2}{2}\left(\cos(\frac{2k-1}{N}\pi) - 1\right), \qquad k = 1, \ldots, \frac{N}{2}. \tag{4.138}$$

These eigenvalues of $\mathbf{A}_1$ (and by similarity of $\mathbf{A}_2$) represent eigenvalues of the complete state matrix $\mathbf{A}$ with algebraic multiplicity 2.

Defining the partial state vectors $\mathbf{x}_1 := [\mathbf{x}]_{1:2:N-1}$ and $\mathbf{x}_2 := [\mathbf{x}]_{2:2:N}$, the system (4.119), (4.121) for $\alpha = \frac{1}{2}$ can be written

$$\begin{aligned} \begin{bmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} u, \\ y &= \begin{bmatrix} \mathbf{c}_1^T & \mathbf{c}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \end{aligned} \tag{4.139}$$

with

$$\mathbf{b}_1 = N \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \frac{1}{4} \end{bmatrix}, \quad \mathbf{b}_2 = N \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \frac{1}{2} \end{bmatrix}, \qquad \mathbf{c}_1^T = N \begin{bmatrix} 2 & 0 & \dots & 0 \end{bmatrix}, \quad \mathbf{c}_2^T = \mathbf{0}^T. \tag{4.140}$$

The subsystem

$$\begin{aligned} \dot{\mathbf{x}}_1 &= \mathbf{A}_1 \mathbf{x}_1 + \mathbf{b}_1 u \\ y &= \mathbf{c}_1^T \mathbf{x}_1 \end{aligned} \tag{4.141}$$

has no invariant zero, i. e. $y(t)$ has relative degree $\frac{N}{2}$ and represents a flat output for this subsystem (the subsystem is controllable and observable). The control $u(t)$ excites the stable internal dynamics (the unobservable subsystem)

$$\dot{\mathbf{x}}_2 = \mathbf{A}_2 \mathbf{x} + \mathbf{b}_2 u, \tag{4.142}$$

whose eigenvalues (i. e. one half of the eigenvalues of $\mathbf{A}$) coincide with the invariant zeros.

A (flatness-based) feedforward control $u(t)$ can be computed based on the inversion of the transfer function $G_1(s) = \mathbf{c}_1^T (s\mathbf{I} - \mathbf{A})^{-1} \mathbf{b}_1$ for the first subsystem, taking into account $y(t)$ and its time derivatives up to order $\frac{N}{2}$, see Chapter 6.

Convergence of the Eigenvalues

We replace the cosines in the discretized eigenvalue expressions (4.133) and (4.138) with their series expansions and obtain for $\alpha = 0$ (see also [111] for this case)

$$\lambda_k = \lambda_{k,\infty} \left(1 - \frac{1}{N} + o\big(\frac{1}{N}\big) \right) \tag{4.143}$$

and for $\alpha = \frac{1}{2}$

$$\lambda_k = \lambda_{k,\infty} \left(1 + \frac{2\lambda_{k,\infty}}{4! \left(\frac{N}{2}\right)^2} + o\Big(\frac{1}{\left(\frac{N}{2}\right)^2}\Big) \right). \tag{4.144}$$

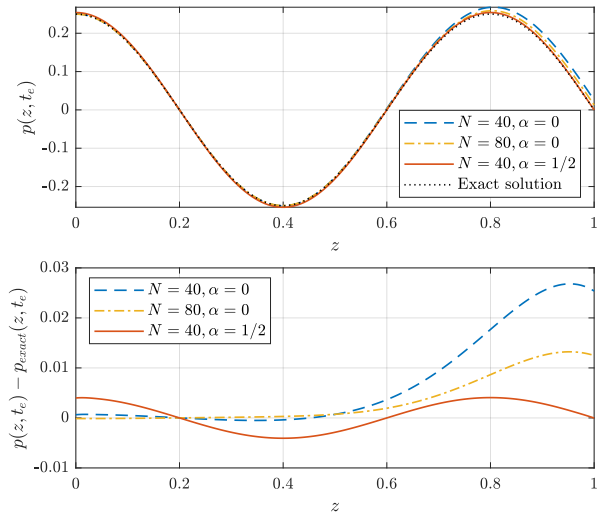

Figure 4.8: Top: Exact and numerical solution of the initial value problem for the heat equation with different parameters and grid sizes. Bottom: Error between numerical and exact solution.

We conclude that the eigenvalues of the heat equation are approximated with a first order error for $\alpha = 0$ and an error of second order for $\alpha = \frac{1}{2}$.

Remark 4.9. The result that $\alpha = \frac{1}{2}$ provides a superior approximation of the eigenvalues supports the observation that a comparable parametrization of the discretization according to [71] (also with a parameter value of $\frac{1}{2}$) yields excellent results for diffusive systems, see [10]. Note however that in [71], the mappings of the nodal efforts are parametrized, while this is the case for the edge flows in our approach. This gives triangular instead of banded matrices $\mathbf{J}_p = -\mathbf{J}_q^T$, which hampers a simple representation of the discretized eigenvalues and their subsequent convergence analysis for $N \to \infty$.

4.6.6.4 Initial Value Problem

In Fig. 4.8, we compare the numerical solution of the heat equation under Neumann-Dirichlet boundary conditions (4.126) with initial condition (only third mode)

$$x_0(t) = x(z,0) = \cos(\frac{5}{2}\pi z), \tag{4.145}$$

with the exact solution at time $t_e = \ln 4 / \left(\frac{5}{2}\pi\right)^2$,

$$x(z,t_e) = e^{-\frac{25}{4}\pi^2 t_e} x_0(z) = \frac{1}{4} x_0(z). \tag{4.146}$$

For $\alpha = 0$, the error over z is reduced by a factor 2 when doubling the number of discretization intervals from $N = 40$ to $N = 80$, which corresponds to the first order approximation of the eigenvalues, see Eq. (4.143). We note that the approximation with $\alpha = \frac{1}{2}$ and $N = 40$ is superior in terms of error magnitude and the shape of the error, which resembles the considered third mode. The results are not surprising, as the *centered* approximation by the parameter choice $\alpha = \frac{1}{2}$ reflects the symmetric character of a diffusive process like heat conduction.

In Chapter 6, we reconsider the finite-dimensional approximate models of the heat equation for the numerical computation of flatness-based feedforward controllers.

4.7 Two-Dimensional Wave Equation

In this section, a special focus is set on the construction and graphical interpretation of the power-preserving mappings. Using Whitney forms, this is facilitated by the meanings of discrete effort and flow degrees of freedom as approximate (integral) quantities on the nodes, edges and faces of the discretization mesh. We present the consistent discretization of the constitutive equations, i. e. we show how to compute the co-state variables on given grid nodes or edges based on the weighted sum of neighboring conserved quantities. We discuss the effects of different parametrizations based on numerical experiments, and close the section with a simulation example on a non-trivial spatial domain.

To illustrate the steps towards an approximate PH state space model with desired boundary inputs by *geometric discretization*, we consider the normalized linear wave equation on a 2-dimensional rectangular domain $\Omega = (0, L_x) \times (0, L_y) \subset \mathbb{R}^2$, with boundary $\partial\Omega$. The spatial domain is covered by a regular, *oriented* simplicial triangulation $\mathcal{T}_h$, as sketched in Fig. 4.9. We recall the structured representation of the wave equation from Subsection 2.3.1 for $n = p = 2$:

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & -\mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix}, \qquad \text{(Structure)} \tag{4.147a}$$

$$\begin{bmatrix} \partial_t p \\ \partial_t q \end{bmatrix} = \begin{bmatrix} -f^p \\ -f^q \end{bmatrix}, \qquad \text{(Dynamics)} \tag{4.147b}$$

$$\begin{bmatrix} e^p \\ e^q \end{bmatrix} = \begin{bmatrix} \delta_p H \\ \delta_q H \end{bmatrix}. \qquad \text{(Constit. Eq.)} \tag{4.147c}$$

The state, flow and effort differential forms are $p, f^p \in L^2\Lambda^2(\Omega)$, $q, f^q \in L^2\Lambda^1(\Omega)$, $e^p \in H^1\Lambda^0(\Omega)$ and $e^q \in H^1\Lambda^1(\Omega)$, and we consider the normalized

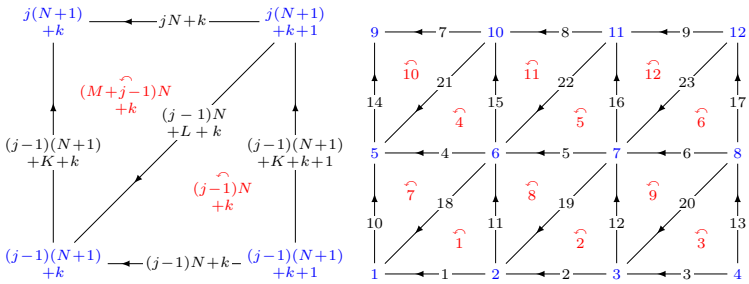

Figure 4.9: Numbering of nodes, edges and faces on a $N \times M$ rectangular simplicial mesh. $K = (M+1)N$, $L = K + M(N+1)$. Left: Unit square cell. Right: 3×2 mesh.

quadratic Hamiltonian functional[21]

$$H = \int_\Omega \frac{1}{2} p \wedge *p + \frac{1}{2} q \wedge *q. \tag{4.148}$$

The boundary input variables (the causality of the boundary ports) will be specified in the *discrete* setting by the choice of the boundary trace matrices $\mathbf{T}_q$ and $\hat{\mathbf{T}}_p$.

4.7.1 Mesh, Matrices and Dimensions

Using Whitney basis forms, the degrees of freedom in the mixed Galerkin approach are associated to integrals of distributed quantities on the k-simplices of the mesh. The dimensions of the (initial) discrete flow and effort vectors equal the numbers of corresponding nodes, edges and faces on the grid. The same holds for the discrete efforts on the boundary, which are designated in- or outputs and are localized on the corresponding boundary nodes and edges, see Table 4.3.

The mixed Galerkin approximation of the Stokes-Dirac structure yields a set of matrices with different sizes and ranks, see Table 4.4. The construction

Table 4.3: Dimensions of discrete flow and efforts spaces on the rectangular $N \times M$ simplicial grid. $\mathbf{e}_b^p$ and $\mathbf{e}_b^q$ contain the corresponding effort degrees of freedom on the complete boundary.

Vector(s)	Dimension	Symbol(s)
$\mathbf{f}^p$	$2NM$	N_p
$\mathbf{f}^q$, $\mathbf{e}^q$	$3NM + N + M$	$N_q = M_q$
$\mathbf{e}^p$	$(N+1)(M+1)$	M_p
$\mathbf{e}_b^p$, $\mathbf{e}_b^q$	$2(N+M)$	$M_p^b = M_q^b$

[21] Which corresponds to a speed of propagation $c = 1$.

Table 4.4: Sizes and ranks of the matrices resulting from the mixed Galerkin approximation and the direct discrete model, respectively. $N, M > 2$.

Matrix	Size	Rank
$\mathbf{M}_p$	$M_p \times N_p$	$M_p - 2$
$\mathbf{K}_p + \mathbf{L}_p$	$M_p \times N_q$	$M_p - 2$
$\mathbf{d}_p$	$N_p \times N_q$	N_p
$\mathbf{M}_q$	$N_q \times N_q$	$2(M_p - 2)$
$\mathbf{K}_q + \mathbf{L}_q$	$N_q \times M_p$	$M_p - 1$
$\mathbf{d}_q$	$N_q \times M_p$	$M_p - 1$
$\mathbf{L}_p = \mathbf{L}_q^T$	$M_p \times N_q$	$2(M + N) - 1$

of power-preserving mappings and conjugated output matrices that satisfy the matrix equation (4.53), is based on rank considerations of the involved matrix products.

4.7.2 Power-Preserving Mappings, Discrete In- and Outputs

We illustrate on three elementary examples the construction of the power-preserving flow and effort maps and conjugated output matrices. The structure of the resulting matrices can be extrapolated to the case of $N \times M$ grids with arbitrarily distributed boundary inputs of mixed causality.

Example 4.3 (Elementary 1×1 grid)**.** Consider the sample grid in Fig. 4.10. The mixed Galerkin discretization of (4.147) with Whitney forms yields the discrete representation (4.43) with $(-1)^r = -1$. The face degrees of freedom (flows) $\mathbf{f}^p = -\dot{\mathbf{p}} \in \mathbb{R}^2$, the edge degrees of freedom (flows and efforts) $\dot{\mathbf{q}} = -\mathbf{f}^q, \mathbf{e}^q \in \mathbb{R}^5$ and the node degrees of freedom (efforts) $\mathbf{e}^p \in \mathbb{R}^4$ are collected in the corresponding vectors. The discrete derivative matrices, which satisfy the discrete complex property $\mathbf{d}_p\mathbf{d}_q = \mathbf{0}$, are the co-incidence matrices of the oriented graph

$$\mathbf{d}_p = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & -1 \end{bmatrix}, \qquad \mathbf{d}_q = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 \end{bmatrix}. \tag{4.149}$$

Input trace matrices and effort maps. We assign all effort degrees of freedom at the boundary edges the role of inputs[22], summarized in $\mathbf{e}^b \in \mathbb{R}^4$.

[22] With this choice, we can easily derive the construction of the power-preserving mappings and output matrices. The same power-preserving mappings are valid with arbitrary boundary causality, while the output matrices can be easily adapted, as in the case of the simulation examples.

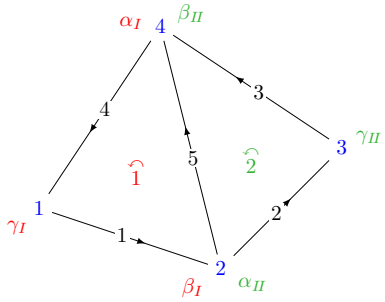

Figure 4.10: Sample grid of 4 nodes to illustrate the construction of power-preserving mappings. The coefficients $\alpha_j, \beta_j, \gamma_j = 1 - \alpha_j - \beta_j$, $j \in \{I, II\}$ weight the contribution of the integral conserved quantities p_1 and p_2 in the definition of the states $\tilde{p}_i$, which are associated to the node efforts (co-states) $\tilde{e}_i^p$, $i = 1, \ldots, 4$.

The interior edge is related to the minimal effort $\tilde{e}^q \in \mathbb{R}$:

$$\begin{bmatrix} \mathbf{e}^b \\ \tilde{e}^q \end{bmatrix} = \begin{bmatrix} \mathbf{T}_q \\ \mathbf{P}_{eq} \end{bmatrix} \mathbf{e}^q \qquad \text{with} \qquad \mathbf{T}_q = \begin{bmatrix} \mathbf{I}_4 & \mathbf{0}_{4\times 1} \end{bmatrix}, \quad \mathbf{P}_{eq} = \begin{bmatrix} \mathbf{0}_{1\times 4} & 1 \end{bmatrix}. \tag{4.150}$$

No node plays the role of an input node, hence,

$$\tilde{\mathbf{e}}^p = \mathbf{P}_{ep}\mathbf{e}^p \qquad \text{with} \qquad \mathbf{P}_{ep} = \mathbf{I}_4. \tag{4.151}$$

Mapping of the conserved quantities on the faces. For the mapping of the vector of integral conserved quantities[23] $\mathbf{p} \in \mathbb{R}^2$ on the two faces (triangles), we argue as follows. The vector of discrete states $\tilde{\mathbf{p}} \in \mathbb{R}^4$, which is dual to the vector $\tilde{\mathbf{e}}^p \in \mathbb{R}^4$ of node efforts, shall

1. contain weighted sums of the discrete conserved quantities on the faces that touch the corresponding node and
2. the sum of its elements must reflect the total conserved quantity. In the example according to Fig. 4.10, this means

$$\sum_{i=1}^{4} \tilde{p}_i = \sum_{j=1}^{2} p_j. \tag{4.152}$$

With $\tilde{\mathbf{p}} = \mathbf{P}_{fp}\mathbf{p}$, the second condition translates to[24] $\mathbf{1}_4^T \mathbf{P}_{fp} = \mathbf{1}_2^T$, i. e. the column sums of the matrix $\mathbf{P}_{fp} \in \mathbb{R}^{4\times 2}$ must equal one. A matrix that satisfies

[23] We refer to the "original" discrete vectors $\mathbf{p}$, $\mathbf{q}$ as *discrete conserved quantities*, while we call $\tilde{\mathbf{p}}$, $\tilde{\mathbf{q}}$ the *state vectors* of the resulting PH state space model.

[24] $\mathbf{1}_n \in \mathbb{R}^n$ denotes a column vector whose n elements are all 1.

this condition is

$$\mathbf{P}_{fp} = \begin{bmatrix} \gamma_I & 0 \\ \beta_I & \alpha_{II} \\ 0 & \gamma_{II} \\ \alpha_I & \beta_{II} \end{bmatrix} \qquad \text{with} \qquad \alpha_j + \beta_j + \gamma_j = 1, \quad j \in \{I, II\}. \tag{4.153}$$

The weights of the conserved quantities p_1, p_2 in the definition of the states $\tilde{p}_i$, which are associated to the nodal efforts $\tilde{e}_i^p$, $i = 1, \ldots, 4$, are printed in Fig. 4.10 in red and green, respectively.

Output matrix for the nodal efforts. The matrix equation (4.53) without a matrix $\hat{\mathbf{T}}_p$ can be written in the form

$$(-1)^r \mathbf{d}_p^T \mathbf{P}_{fp}^T \mathbf{P}_{ep} + \begin{bmatrix} \mathbf{T}_q^T & \mathbf{P}_{eq}^T \end{bmatrix} \begin{bmatrix} \mathbf{S}_p \\ \mathbf{P}_{fq}\mathbf{d}_q \end{bmatrix} = \mathbf{0}. \tag{4.154}$$

Exploiting that $\begin{bmatrix} \mathbf{T}_q^T & \mathbf{P}_{eq}^T \end{bmatrix}$ is a permutation matrix, the equation can be multiplied from the left with its transpose (which equals its inverse), and we obtain as the first line the output matrix associated to node efforts

$$\mathbf{S}_p = -(-1)^r \mathbf{T}_q \mathbf{d}_p^T \mathbf{P}_{fp}^T \mathbf{P}_{ep} = \begin{bmatrix} \gamma_I & \beta_I & 0 & \alpha_I \\ 0 & \alpha_{II} & \gamma_{II} & \beta_{II} \\ 0 & \alpha_{II} & \gamma_{II} & \beta_{II} \\ \gamma_I & \beta_I & 0 & \alpha_I \end{bmatrix}. \tag{4.155}$$

The discrete output vector $\mathbf{f}^b = \mathbf{S}_p \mathbf{e}^p$ contains – on this very simple grid – two pairs of identical elements, which each represent convex sums of the node efforts. Regarding for example the outer boundary of face 1 in Fig. 4.10, this identity is no surprise. If we delete node 1 (from the graph), and consider edges 1 and 4 as a single edge 14, the power $e_1^b f_1^b + e_4^b f_4^b$ which is transmitted over both edges must equal $(e_1^b + e_4^b) f_{14}^b$, which is the case for $f_1^b = f_4^b$.

Mapping of the edge states. In analogy to (4.155), the matrix equation

$$\mathbf{P}_{fq}\mathbf{d}_q = -(-1)^r \mathbf{P}_{eq} \mathbf{d}_p^T \mathbf{P}_{fp}^T \mathbf{P}_{ep} \tag{4.156}$$

determines the matrix $\mathbf{P}_{fq}$. The solution consists of a particular part to which a linear combination of the rows of $\mathbf{d}_p$ (recall that $\mathbf{d}_p \mathbf{d}_q = \mathbf{0}$) can be added:

$$\begin{aligned} \mathbf{P}_{fq} &= \mathbf{P}_{fq}^p + \begin{bmatrix} c_1 & c_2 \end{bmatrix} \mathbf{d}_p \\ &= \begin{bmatrix} -\gamma_I & -\gamma_{II} & \alpha_I - \beta_{II} & 0 & 0 \end{bmatrix} + \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & -1 \end{bmatrix}. \end{aligned} \tag{4.157}$$

With $c_1 = \frac{\gamma_I}{2}$ and $c_2 = -\alpha_I + \beta_{II} + \frac{\gamma_{II}}{2}$, we get a matrix of the form

$$\begin{aligned} \mathbf{P}_{fq} &= \mathbf{P}_{fq}^{\perp} + \mathbf{P}_{fq}^{\parallel} \\ &= \begin{bmatrix} -\frac{\gamma_I}{2} & -\frac{\gamma_{II}}{2} & \frac{\gamma_{II}}{2} & \frac{\gamma_I}{2} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{\alpha_I - \beta_I}{2} + \frac{\alpha_{II} - \beta_{II}}{2} \end{bmatrix}, \end{aligned} \tag{4.158}$$

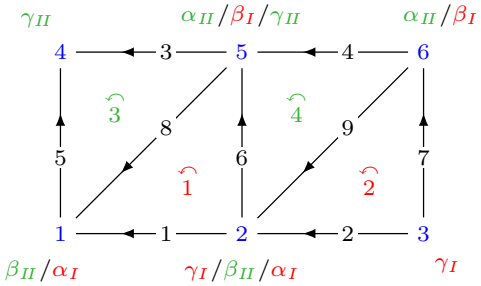

Figure 4.11: Illustration of the 2×1 grid in Example 4.4. Each state $\tilde{p}_i$, which is associated to a "nodal" co-state $\tilde{e}_i^p$, $i = 1, \ldots, 6$, is defined by a weighted sum of the conserved quantities p_j, $j = 1, \ldots, 4$, on the adjacent triangles. The weights are printed next to the nodes. Red color and index I refer to the lower triangles, green color and index II to the upper triangles.

where $\mathbf{P}_{fq}^{\perp}$ contains the weights of the conserved quantities q_j on the edges "across" the edge on which the minimal effort $\tilde{e}^q$ is defined. Accordingly, $\mathbf{P}_{fq}^{\parallel}$ contains the weight associated to exactly this edge. Note that only $\mathbf{P}_{fq}^{\perp}$ will contribute to the definition of the discrete Hodge matrix $\mathbf{Q}_q$, which relates the efforts *across* edges of the grid with the states *along the dual edges*, see Fig. 4.12 in Example 4.5.

The construction, which we demonstrated for the simplest quadrilateral grid, can be extended to a rectangular grid, which is shown in the next example.

Example 4.4 (2×1 grid, unique boundary causality)**.** We now consider the 2×1 rectangular grid as depicted in Fig. 4.11, whose co-incidence matrices are the discrete derivative matrices

$$\mathbf{d}_p = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 \end{bmatrix},$$

$$\mathbf{d}_q = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}. \tag{4.159}$$

Input trace matrices and effort mappings. As in the previous example, we start with a single causality on the boundary and the only input trace matrix

$$\mathbf{T}_{q,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{4.160}$$

The remaining edges and all nodes are the discrete objects on which the elements of the co-state vectors $\tilde{\mathbf{e}}^q$ and $\tilde{\mathbf{e}}^p$ are defined. This fact is represented by the effort mapping matrices

$$\mathbf{P}_{eq,1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{P}_{ep,1} = \mathbf{I}_6. \tag{4.161}$$

We use the index 1 for this case with *only edge inputs*, and refer to the corresponding matrices in the following example.

Mapping of the conserved quantities on the faces. With the same arguments as for the simple example before, we can construct the matrix to define the discrete states $\tilde{\mathbf{p}} = \mathbf{P}_{fp,1}\mathbf{p}$, see also the illustration of the weights in Fig. 4.11:

$$\mathbf{P}_{fp,1} = \begin{bmatrix} \alpha_I & 0 & \beta_{II} & 0 \\ \gamma_I & \alpha_I & 0 & \beta_{II} \\ 0 & \gamma_I & 0 & 0 \\ 0 & 0 & \gamma_{II} & 0 \\ \beta_I & 0 & \alpha_{II} & \gamma_{II} \\ 0 & \beta_I & 0 & \alpha_{II} \end{bmatrix}, \qquad \alpha_{I/II} + \beta_{I/II} + \gamma_{I/II} = 1. \tag{4.162}$$

Output matrices for the nodal efforts. According to (4.155) we obtain for the nodal output matrix

$$\mathbf{S}_{p,1} = \begin{bmatrix} -\alpha_I & -\gamma_I & 0 & 0 & -\beta_I & 0 \\ 0 & -\alpha_I & -\gamma_I & 0 & 0 & -\beta_I \\ \beta_{II} & 0 & 0 & \gamma_{II} & \alpha_{II} & 0 \\ 0 & \beta_{II} & 0 & 0 & \gamma_{II} & \alpha_{II} \\ -\beta_{II} & 0 & 0 & -\gamma_{II} & -\alpha_{II} & 0 \\ 0 & \alpha_I & \gamma_I & 0 & 0 & \beta_I \end{bmatrix}. \tag{4.163}$$

Note that again there are two pairs of identical outputs (modulo the sign depending on the orientation of the input edge), which is due to the fact that by merging the adjacent edges, nodes 3 and 4 could be removed from the graph.

Mapping of the edge states. The solution of the matrix equation (4.156) for the matrices as defined above (again, the rows of $\mathbf{d}_p$ can be used to adjust the solution) results in a matrix

$$\mathbf{P}_{fq,1} = \mathbf{P}^{\perp}_{fq,1} + \mathbf{P}^{\parallel}_{fq,1} + \mathbf{P}^{rot}_{fq,1} \tag{4.164}$$

with

$$\mathbf{P}^{\perp}_{fq,1} = \left[\begin{array}{ccccccc|cc} \alpha_I & 0 & 0 & \alpha_{II} & 0 & 0 & 0 & 0 & 0 \\ -\frac{\gamma_I}{2} & 0 & -\frac{\gamma_{II}}{2} & 0 & -\frac{\gamma_{II}}{2} & -\frac{\gamma_I}{2} & 0 & 0 & 0 \\ 0 & -\frac{\gamma_I}{2} & 0 & -\frac{\gamma_{II}}{2} & 0 & -\frac{\gamma_{II}}{2} & -\frac{\gamma_I}{2} & 0 & 0 \end{array}\right], \tag{4.165a}$$

$$\mathbf{P}^{\parallel}_{fq,1} = \left[\begin{array}{ccccccc|cc} 0 & 0 & 0 & 0 & 0 & \beta_I+\beta_{II}-1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\alpha_I-\beta_I}{2}+\frac{\alpha_{II}-\beta_{II}}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\alpha_I-\beta_I}{2}+\frac{\alpha_{II}-\beta_{II}}{2} \end{array}\right], \tag{4.165b}$$

$$\mathbf{P}^{rot}_{fq,1} = \left[\begin{array}{ccccccc|cc} -\delta_I & \delta_{II} & \delta_I & -\delta_{II} & -\delta_I & \delta_I+\delta_{II} & -\delta_{II} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \tag{4.165c}$$

and the abbreviation

$$\delta_{I/II} = \frac{1}{8} + \frac{1}{4}(\alpha_{I/II} - \beta_{I/II}). \tag{4.166}$$

Note that the definition of the state vector $\tilde{\mathbf{q}} = \mathbf{P}_{fq,1}\mathbf{q}$ now contains a *rotational* component, which is illustrated in Fig. 4.12 (for the following example).

Example 4.5 (2×1 grid, mixed boundary causality)**.** Still considering the grid in Fig. 4.11, we assign the efforts in nodes 1 and 2 the role of (boundary) inputs $\hat{e}^b_1$ and $\hat{e}^b_2$ and remove the effort on edge 1 from the input vector $\mathbf{e}^b$. The corresponding input trace matrices are

$$\hat{\mathbf{T}}_p = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{T}_q = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{4.167}$$

and the effort mappings

$$\mathbf{P}_{ep} = \begin{bmatrix} \mathbf{0}_{4\times 2} & \mathbf{I}_4 \end{bmatrix}, \qquad \mathbf{P}_{eq} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.168}$$

The matrix equation (4.53) for power preservation can now be written as

$$\begin{bmatrix} (-1)^r \mathbf{d}_p^T \mathbf{P}_{fp}^T & \hat{\mathbf{S}}_q^T \end{bmatrix} \begin{bmatrix} \mathbf{P}_{ep} \\ \hat{\mathbf{T}}_p \end{bmatrix} + \begin{bmatrix} \mathbf{P}_{eq}^T & \mathbf{T}_q^T \end{bmatrix} \begin{bmatrix} \mathbf{P}_{fq}\mathbf{d}_q \\ \mathbf{S}_p \end{bmatrix} = \mathbf{0}. \tag{4.169}$$

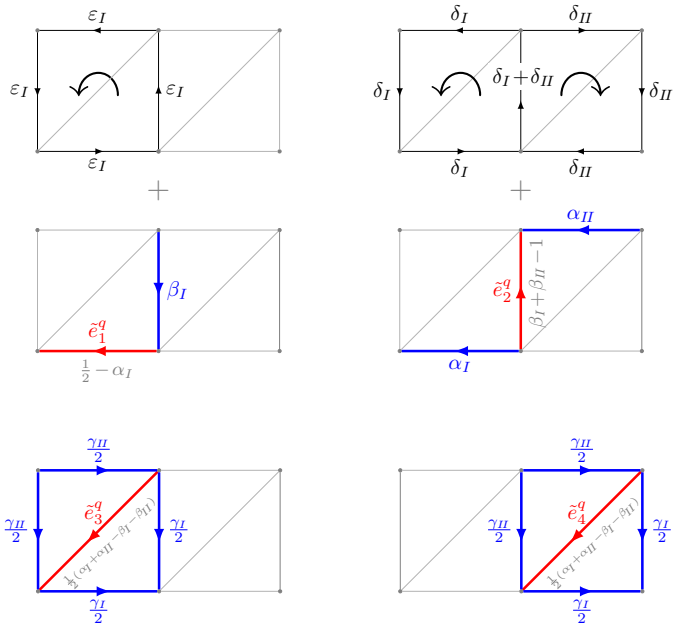

Figure 4.12: Illustration of the components of $\tilde{\mathbf{q}}$ ($\tilde{q}_1$ to $\tilde{q}_4$ from upper left to lower right) in Example 4.5 in terms of the original conserved quantities q_j, $j = 1, \ldots, 9$, on the edges of the grid. The components across the considered effort edge/aligned with the effort edge are drawn in blue/red. The round black arrows indicate the sense of the rotational components in $\tilde{q}_1$ and $\tilde{q}_2$ for positive values of ε_I and $\delta_{I/II}$, respectively.

For the moment, we assume that by appropriate choice of $\hat{\mathbf{S}}_q$, the first term can be made $(-1)^r \mathbf{d}_p^T \mathbf{P}_{fp,1}^T \mathbf{P}_{ep,1}$. We obtain the flow map $\mathbf{P}_{fq}$ and the output matrix $\mathbf{S}_p$ in the second term (with $\mathbf{P}_{ep,1} = \mathbf{I}$) by the solution of

$$\mathbf{P}_{fq}\mathbf{d}_q = -(-1)^r \mathbf{P}_{eq}\mathbf{d}_p^T \mathbf{P}_{fp,1}^T, \qquad \mathbf{S}_p = -(-1)^r \mathbf{T}_q \mathbf{d}_p^T \mathbf{P}_{fp,1}^T. \tag{4.170}$$

The output matrix $\mathbf{S}_p$ contains the rows of $\mathbf{S}_{p,1}$ that correspond to the input edges represented by the rows of $\mathbf{T}_q$. In the present case, we have to delete the first row in (4.163) and obtain

$$\mathbf{S}_p = \begin{bmatrix} 0 & -\alpha_I & -\gamma_I & 0 & 0 & -\beta_I \\ \beta_{II} & 0 & 0 & \gamma_{II} & \alpha_{II} & 0 \\ 0 & \beta_{II} & 0 & 0 & \gamma_{II} & \alpha_{II} \\ -\beta_{II} & 0 & 0 & -\gamma_{II} & -\alpha_{II} & 0 \\ 0 & \alpha_I & \gamma_I & 0 & 0 & \beta_I \end{bmatrix}. \tag{4.171}$$

The construction of $\mathbf{P}_{fq}$ follows the same lines as in the previous examples. The horizontal edge, on which a discrete co-state is defined, gives rise to a new element of the discrete state vector $\tilde{\mathbf{q}} \in \mathbb{R}^4$, which is illustrated in Fig. 4.12.

The matrix $\mathbf{P}_{fq}$ becomes

$$\mathbf{P}_{fq} = \mathbf{P}_{fq}^{\perp} + \mathbf{P}_{fq}^{\parallel} + \mathbf{P}_{fq}^{rot} \tag{4.172}$$

with

$$\mathbf{P}_{fq}^{\perp} = \left[\begin{array}{ccccccc|cc} 0 & 0 & 0 & 0 & 0 & -\beta_I & 0 & 0 & 0 \\ \alpha_I & 0 & 0 & \alpha_{II} & 0 & 0 & 0 & 0 & 0 \\ -\frac{\gamma_I}{2} & 0 & -\frac{\gamma_{II}}{2} & 0 & -\frac{\gamma_{II}}{2} & -\frac{\gamma_I}{2} & 0 & 0 & 0 \\ 0 & -\frac{\gamma_I}{2} & 0 & -\frac{\gamma_{II}}{2} & 0 & -\frac{\gamma_{II}}{2} & -\frac{\gamma_I}{2} & 0 & 0 \end{array}\right], \tag{4.173a}$$

$$\mathbf{P}_{fq}^{\parallel} = \left[\begin{array}{ccccccc|cc} \frac{1}{2}-\alpha_I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_I+\beta_{II}-1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\alpha_I-\beta_I}{2}+\frac{\alpha_{II}-\beta_{II}}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\alpha_I-\beta_I}{2}+\frac{\alpha_{II}-\beta_{II}}{2} \end{array}\right], \tag{4.173b}$$

$$\mathbf{P}_{fq}^{rot} = \left[\begin{array}{ccccccc|cc} -\varepsilon_I & 0 & \varepsilon_I & 0 & -\varepsilon_I & \varepsilon_I & 0 & 0 & 0 \\ -\delta_I & \delta_{II} & \delta_I & -\delta_{II} & -\delta_I & \delta_I+\delta_{II} & -\delta_{II} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \tag{4.173c}$$

and the new abbreviation

$$\varepsilon_{I/II} = \frac{1}{8} - \frac{1}{4}(\alpha_{I/II} - \beta_{I/II}). \tag{4.174}$$

Figure 4.12 illustrates the different components whose (vector) sums constitute the states $\tilde{q}_i$, $i = 1, \ldots, 4$, in the example. With

$$\mathbf{P}_{fp} = \mathbf{P}_{ep}\mathbf{P}_{fp,1} = \begin{bmatrix} 0 & \gamma_I & 0 & 0 \\ 0 & 0 & \gamma_{II} & 0 \\ \beta_I & 0 & \alpha_{II} & \gamma_{II} \\ 0 & \beta_I & 0 & \alpha_{II} \end{bmatrix} \tag{4.175}$$

and

$$\begin{aligned} \hat{\mathbf{S}}_q &= (-1)^r \hat{\mathbf{T}}_p \begin{bmatrix} \mathbf{d}_q^T \mathbf{P}_{fq}^T & \mathbf{S}_p^T \end{bmatrix} \begin{bmatrix} \mathbf{P}_{eq} \\ \mathbf{T}_q \end{bmatrix} \\ &= \begin{bmatrix} \alpha_I - \frac{1}{2} & 0 & -\beta_{II} & 0 & \beta_{II} & -\alpha_I & 0 & \beta_{II}-\alpha_I & 0 \\ \gamma_I - \frac{1}{2} & \alpha_I & 0 & -\beta_{II} & 0 & \beta_{II}-\gamma_I & -\alpha_I & -\gamma_I & \beta_{II}-\alpha_I \end{bmatrix}, \end{aligned} \tag{4.176}$$

see Fig. 4.13, the parametrization of power-preserving effort and flow maps and output matrices is completed[25].

[25] It is straightforward to verify that the latter matrices indeed satisfy $(-1)^r \mathbf{d}_p^T \mathbf{P}_{fp}^T \mathbf{P}_{ep} + \hat{\mathbf{S}}_q^T \hat{\mathbf{T}}_p = (-1)^r \mathbf{d}_p^T \mathbf{P}_{fp,1}^T \mathbf{P}_{ep,1}$, which has been assumed at the beginning of the example.

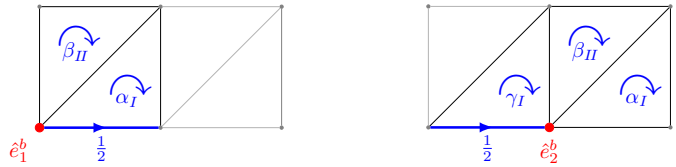

Figure 4.13: Illustration of the geometric objects on which the elements $\hat{f}_1^b$ and $\hat{f}_2^b$ of the output vector $\hat{\mathbf{f}}^b$ are defined in Example 4.5 (blue, with weights). The elements of the (dual) input vector $\hat{\mathbf{e}}^b$ are defined on the red nodes.

4.7.3 Generalization to $N \times M$ Meshes and Remarks

$N \times M$ meshes. The construction as presented on the three elementary examples above can be generalized in a straightforward manner to arbitrary $N \times M$ rectangular meshes. The direct interpretation of the discretized system equations as discrete conservation laws in the case of Whitney approximation forms allows for a construction of the matrices based on the properties of the 2-complex (generalized oriented graph) on the discretization mesh. In the above examples, we used only two sets of convex weights $(\alpha_j, \beta_j, \gamma_j)$, $j \in \{I, II\}$ for the upper and lower triangles. It is, however, possible to assign different combinations of convex weights to each triangle, for example on non-rectangular meshes over more complex geometries.

Input trace matrices and effort maps. Identifying the elements of the input vector $\mathbf{u} = \begin{bmatrix}(\mathbf{u}^p)^T & (\mathbf{u}^q)^T\end{bmatrix}^T = \begin{bmatrix}(\hat{\mathbf{e}}^b)^T & (\mathbf{e}^b)^T\end{bmatrix}^T$ with effort degrees of freedom on the boundary nodes and edges corresponds to a consistent imposition of the effort boundary conditions in the finite-dimensional model. To arrive at the input-output representation (4.58), the matrices

$$\mathbf{\Pi}_p := \begin{bmatrix}\mathbf{P}_{ep} \\ \hat{\mathbf{T}}_p\end{bmatrix} \qquad \text{and} \qquad \mathbf{\Pi}_q := \begin{bmatrix}\mathbf{P}_{eq} \\ \mathbf{T}_q\end{bmatrix} \tag{4.177}$$

should be square and invertible. With the presented choice, $\mathbf{\Pi}_p$ and $\mathbf{\Pi}_q$ become *permutation matrices* and the property $\mathbf{\Pi}_{p/q}^{-1} = \mathbf{\Pi}_{p/q}^T$ makes the matrices of the state space model as indicated in (4.61) particularly simple.

Flow/state maps. By the presented construction, each element $\tilde{p}_i$, $i = 1, \ldots, \tilde{N}_p$, of $\tilde{\mathbf{p}} = \mathbf{P}_{fp}\mathbf{p}$ is related to a 2-chain (a weighted formal sum of 2-simplices), located around the node associated to $\tilde{e}_i^p$. The node and the weighted 2-chain can be considered as *topologically dual* objects. The property $\alpha_\nu + \beta_\nu + \gamma_\nu = 1$, $\nu \in \{I, II\}$, ensures that the balance of the discrete conserved

quantities

$$\sum_{i=1}^{\tilde{N}_p} \tilde{p}_i = \sum_{j=1}^{N_p} p_j - \epsilon_p \tag{4.178}$$

holds. If (boundary) input nodes are defined, the error $\epsilon_p \neq 0$ occurs, because the weighted contribution of p_j on 2-simplices next to the input nodes is neglected in the definition of discrete states. It is easy to imagine that the error ϵ^p, which tends to zero with grid refinement, can be related to well-known effects from the discretization with staggered grids, like ghost values, see e. g. Chapter 3.

A related interpretation of the (minimal) states in terms of topologically dual objects holds for the different elements of the vector $\tilde{\mathbf{q}} = \mathbf{P}_{fq}\mathbf{q}$. As shown in Fig. 4.12, each element $\tilde{q}_i$ of $\tilde{\mathbf{q}}$ can be considered dual to a discrete effort $\tilde{e}_i^q$ on a horizontal, vertical or diagonal edge (drawn in red). $\tilde{q}_i$ is localized on a formal sum of the adjacent 1-simplices (edges), which can be decomposed into components across and along the effort edge and a rotational part. Only the "across" part contributes to the discrete constitutive equations as discussed in the next subsection. While the effort edges are considered *outer oriented* ("across"), the formal sums of edges, on which the $\tilde{q}_i$ are defined, are *inner oriented* ("along"), which describes the geometric nature of the different system variables.

Remark 4.10. The reconstruction of the rotational components of $\tilde{\mathbf{q}}$ from the given quantities can be used to discretize the vorticity term in the shallow water equations (2.73).

Power-conjugated discrete outputs. Like the minimal flows and efforts, the discrete power-conjugated outputs $\mathbf{f}^b = \mathbf{S}_p\mathbf{e}^p$ and $\hat{\mathbf{f}}^b = \hat{\mathbf{S}}_q\mathbf{e}^q$ are constructed as weighted sums of the discrete efforts in the vicinity of the corresponding boundary input. The components f_i^b are defined by a convex sum of node efforts, see e. g. (4.171). The $\hat{f}_i^b$ are composed of rotational parts and a component associated to the neighboring, outer oriented boundary edge, as illustrated in Fig. 4.13.

If the effort maps and input trace matrices form permutation matrices (4.177), the feedthrough matrices in the PH state space model according to (4.58) become $\mathbf{D}_q = \hat{\mathbf{S}}_q\mathbf{T}_q^T$ and $\mathbf{D}_p = \mathbf{S}_p\hat{\mathbf{T}}_p^T$, see Eq. (4.61). By the collocated construction of $\mathbf{f}^b$ and $\hat{\mathbf{f}}^b$, these matrices have only non-zero elements at the interfaces between two boundary regions Γ_i and $\hat{\Gamma}_j$ with different causality. This feedthrough is *physical* as it only stems from the definition of neighboring in- and outputs, and can be completely avoided by setting the boundary inputs zero at these interfaces. For 1D systems, where the two parts of the boundary are not connected, *no feedthrough term* occurs at all. The absence of an undesired direct feedthrough (undesired at least for the numerical approximation of hyperbolic systems) distinguishes our method from the structure-preserving

discretization according to [71], where the feedthrough stems from the convex sum of nodal efforts to define the discrete co-state variables.

4.7.4 Discrete Constitutive Equations

To obtain a consistent numerical approximation of the system of conservation laws, the discrete states $\tilde{\mathbf{p}}$, $\tilde{\mathbf{q}}$ and the efforts or co-states $\tilde{\mathbf{e}}^p$, $\tilde{\mathbf{e}}^q$ must be related via discrete constitutive relations that are consistent with the continuous ones. We consider the case of linear constitutive equations

$$e^p = *p, \qquad e^q = -*q \tag{4.179}$$

that are derived from the Hamiltonian functional over $\Omega \subset \mathbb{R}^2$ with quadratic Hamiltonian density $\mathcal{H} = \frac{1}{2}p\wedge *p + \frac{1}{2}q\wedge *q$, see (4.148). The discrete constitutive equations will be expressed by

$$\tilde{\mathbf{e}}^p = \mathbf{Q}_p\tilde{\mathbf{p}}, \qquad \tilde{\mathbf{e}}^q = \mathbf{Q}_q\tilde{\mathbf{q}} \tag{4.180}$$

with positive definite, diagonal matrices $\mathbf{Q}_p$, $\mathbf{Q}_q$ that represent *diagonal discrete Hodge operators* [178]. The discrete states $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$ are constructed (as $\tilde{\mathbf{f}}^p$ and $\tilde{\mathbf{f}}^q$) as linear combinations of integral conserved quantities on the 2- and 1-simplices of the discretization grid. The faces, based on which $\tilde{p}_i$ is constructed, surround the node to which $\tilde{e}_i^p$ is associated. A similar *geometric* duality[26] can be observed for the $\tilde{e}_i^q$-edges and the neighboring edges that constitute $\tilde{q}_i$. One can even imagine $\tilde{\mathbf{p}}$, $\tilde{\mathbf{q}}$ localized on a (virtual) *dual* grid, whose localization and shift with respect to the original (primal) grid are *parametrized* by the convex set of mapping parameters $(\alpha_j, \beta_j, \gamma_j)$, which we assume all to be positive and related via $\alpha_j + \beta_j + \gamma_j = 1$. Moreover, we consider a mesh with equal step size $h_x = h_y = h$ in both coordinate directions ($h_x = \frac{L_x}{N}$, $h_y = \frac{L_y}{M}$).

For the consistent discretization of the time-invariant constitutive equations, we consider a steady state. In this case, the elements of $\tilde{\mathbf{e}}^p$ must represent "average" values of p on the *weighted* sum of balance areas[27] on which the states $\tilde{p}_i$ are defined. The diagonal matrix $\mathbf{Q}_p$ with elements

$$[\mathbf{Q}_p]_{i,i} = \frac{2}{h^2 \sum_{j=1}^{N_p} [\mathbf{P}_{fp}]_{i,j}}, \qquad i = 1, \ldots, \tilde{N}_p, \tag{4.181}$$

represents a consistent Hodge matrix.

Accordingly, the elements of $\tilde{\mathbf{e}}^q$ must reflect the integral flux of the vector field[28] $q^\sharp$ across the corresponding horizontal, vertical or diagonal edges. Only the parts of $\tilde{q}_i$, which are associated to the edges *perpendicular* to the $\tilde{e}_i^q$-edge,

[26] This geometric duality is immediately given if the two conservation laws are modeled on *two* shifted grids, i. e. dual meshes [174].

[27] Precisely, the average value of the coefficient function of the 2-form p.

[28] Index raising of the 1-form q.

contribute to this flux. This reasoning yields a *diagonal* matrix $\mathbf{Q}_q$ that replaces the Hodge star in (4.148) with diagonal elements[29]

$$[\mathbf{Q}_q]_{i,i}^{hor/ver} = \frac{1}{\sum_{j=1}^{N_q} \left|[\mathbf{P}_{fq}^{\perp}]_{i,j}\right|} \qquad \text{and} \qquad [\mathbf{Q}_q]_{i,i}^{dia} = \frac{2}{\sum_{j=1}^{N_q} \left|[\mathbf{P}_{fq}^{\perp}]_{i,j}\right|} \tag{4.182}$$

for the efforts on horizontal/vertical and diagonal edges, respectively. The absolute values stem from the fact that the *orientation* of a state $\tilde{q}_i$ is coded by the sign of the corresponding row of $\mathbf{P}_{fq}$.

4.7.5 Simulation: Wave Propagation on a Square

We consider the linear wave equation in port-Hamiltonian form (4.147), (4.148) on a square domain $\Omega = (0, 20) \times (0, 20)$ to illustrate the effects of different mapping parameters. We impose the boundary conditions

$$\begin{aligned} e^p(0, 0, t) = u(t) &= \begin{cases} \sin^2(\frac{\pi}{8}t), & 0 \le t < 8, \\ 0, & t \ge 8, \end{cases} \\ e^q(x, y, t) &= 0 \quad \text{on} \quad \partial\Omega \end{aligned} \tag{4.183}$$

by the input trace matrices

$$\hat{\mathbf{T}}_p = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}, \qquad \mathbf{T}_q = \mathbb{I}_b^1, \tag{4.184}$$

where $\mathbb{I}_b^1 \in \mathbb{R}^{M_q^b \times N_q}$ is the matrix composed of unit row vectors associated to the boundary edges. The inputs to the simulation model according to Eq. (4.64) are

$$\hat{e}^b(t) = u(t), \qquad \mathbf{e}^b(t) = \mathbf{0}. \tag{4.185}$$

Table 4.5: Parameter sets used in the simulations.

	#1	#2	#3	#4
α_I	1/3	1/2	2/3	15/16
β_I	1/3	1/4	1/12	1/32
γ_I	1/3	1/4	1/4	1/32
δ_I	1/8	3/16	13/48	45/128
ε_I	1/8	1/16	−1/48	−13/128
α_{II}	1/3	1/4	1/12	1/32
β_{II}	1/3	1/2	2/3	15/16
γ_{II}	1/3	1/4	1/4	1/32
δ_{II}	1/8	1/16	−1/48	−13/128
ε_{II}	1/8	3/16	13/48	45/128

[29] Note that our grids according to Fig. 4.9 have square cells and unique orientations of horizontal, vertical and diagonal edges.

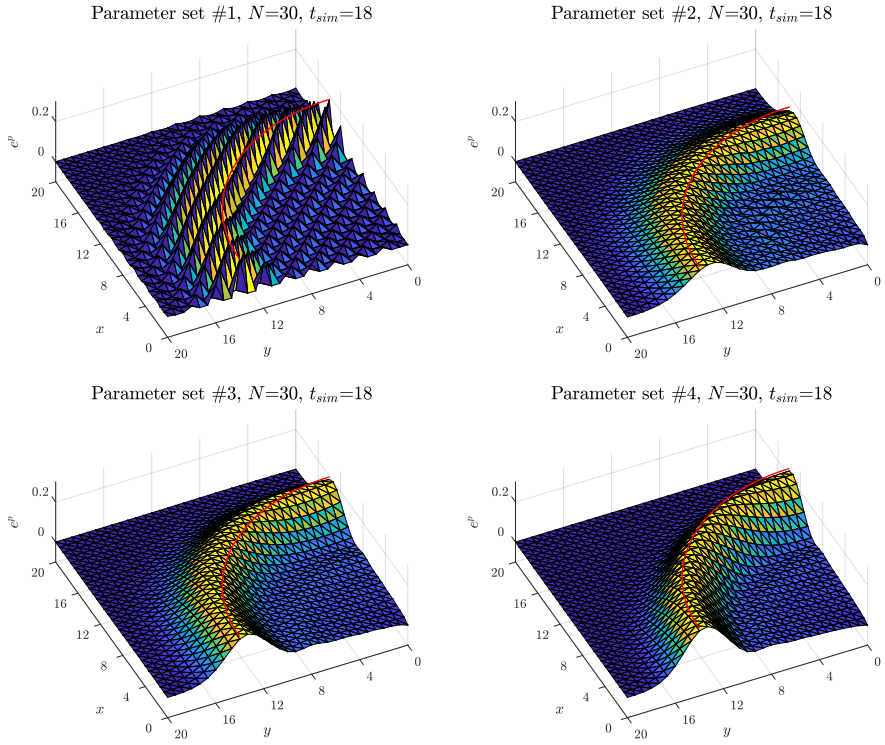

Figure 4.14: Propagation of a wave due to point-wise boundary excitation under different parametrizations of the method. Snapshots at $t_{sim} = 18$.

Fig. 4.14 shows the simulated propagation of the wave in radial direction under different parametrizations of the method. The red line displays a circle with radius $t_{sim} - T/2 = 14$, as a reference for the maximum of the wave front[30] at time t_{sim}, based on the exact solution. The parameter sets in Table 4.5 represent different weightings of the 2-simplices in the propagation direction to compute $\tilde{f}_i^p$, see Fig. 4.15. For parameter set #1 (equal weights 1/3 in the definition of $\tilde{f}_i^p$ associated to a nodal effort $\tilde{e}_i^p$), the propagation of the wave front in the effort variable e^p is reproduced in a completely unsatisfactory manner. Parameter set #2 leads also to undesired dispersion. Moreover, the quarter circle shape of the wave is perturbed, which is due to the non-isotropic mesh and the inadequate parametrization. Parameter set #3 shows less dispersion and parametrization #4 reproduces appropriately the circular wave front

A direct explanation of the unsatisfactory behavior of the numerical solutions #1 and #2 can be found by studying the definition of the matrix $\mathbf{P}_{fq}^{\perp}$,

[30] The plots in Fig. 4.14 represent the discrete, minimal efforts $\tilde{e}_i^p$ *in the nodes* of the mesh.

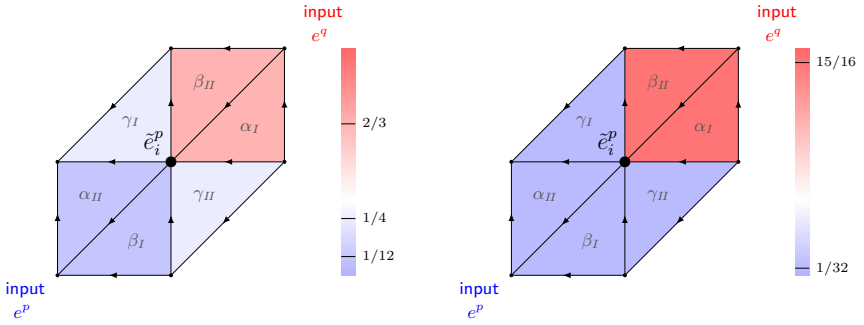

Figure 4.15: Illustration of 2-simplices and weights that contribute to the definition of the discrete state $\tilde{p}_i$ for the parametrizations # 3 (left) and #4 (right). Note that information of the conserved quantity p, which is directly influenced by the boundary input effort e^q in the upper right corner, is preferred for the computation of the node effort $\tilde{e}_i^p$ ("upwinding").

which is visualized in the upper drawings of Fig. 4.12 for the elementary example. Consider first the parametrizations #3 and #4 in Table 4.5. With

$$\operatorname{sgn}(\delta_I) = -\operatorname{sgn}(\delta_{II}) \qquad \text{and} \qquad \operatorname{sgn}(\varepsilon_I) = -\operatorname{sgn}(\varepsilon_{II}), \tag{4.186}$$

the rotational parts in the definition of the discrete states $\tilde{\mathbf{q}}$ are composed of discrete rotations of $\mathbf{q}$ *in the same sense*. This is not the case for parametrizations #1 and #2, which is a hint that reasonable parameter sets for the numerical approximation of hyperbolic systems should satisfy condition (4.186), or, equivalently, $\alpha_I - \beta_I \leqslant \frac{1}{2}$ and at the same time $\alpha_{II} - \beta_{II} \geqslant \frac{1}{2}$. Note that all four simulation models are *conservative by construction* due to the preservation of the port-Hamiltonian structure.

4.7.6 Simulation: Double Slit Experiment

To illustrate the applicability of our approach to more complex spatial domains, we consider the double slit experiment, as exposed in [110], Section 5.8. The linear wave equation with speed of propagation one is considered on a square domain $(0,1) \times (0,1)$, which is complemented by two narrow strips on $(\frac{1}{3}, \frac{5}{12}) \times (-\frac{1}{4}, 0)$ and $(\frac{7}{12}, \frac{2}{3}) \times (-\frac{1}{4}, 0)$. The boundary of the spatial domain Ω is composed of the "lower ends" of the strips

$$\hat{\Gamma} = \{(x,y) \in \partial\Omega \mid y = -\tfrac{1}{4}\} \tag{4.187}$$

and the complement $\Gamma = \partial\Omega \backslash \hat{\Gamma}$. The Dirichlet boundary condition

$$e^p(x,y,t) = u(t) = 0.1 \sin(8\pi t), \qquad (x,y) \in \hat{\Gamma}, \quad t \geq 0, \tag{4.188}$$

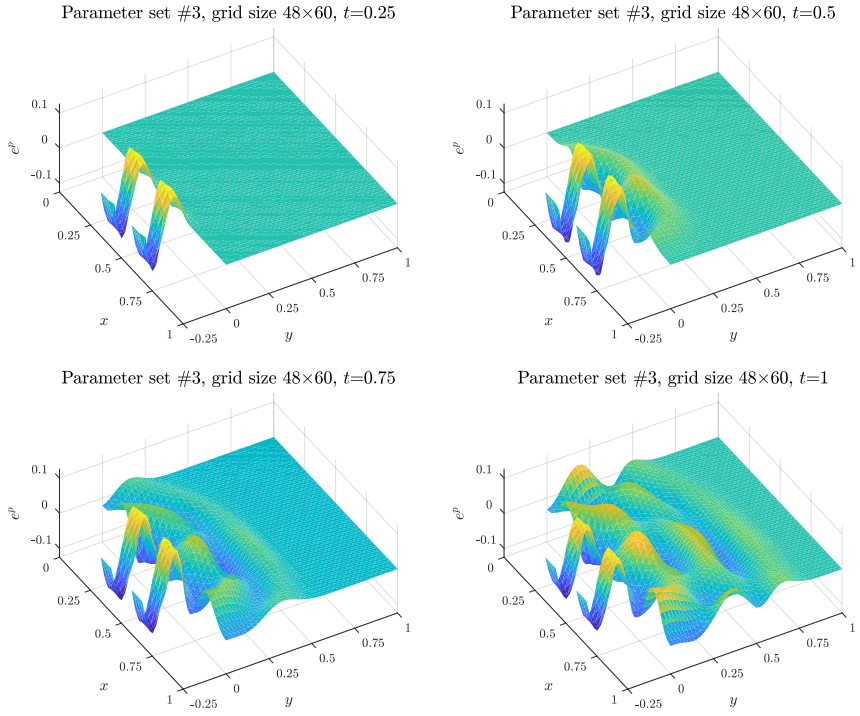

Figure 4.16: Snapshots of the double slit experiment with parameter set #3.

serves as a time varying input on $\hat{\Gamma}$, while the rest of the boundary is subject to the homogeneous Neumann boundary condition

$$e^q(x, y, t) = 0, \qquad (x, y) \in \Gamma, \quad t \geq 0. \tag{4.189}$$

The matrices for the finite-dimensional PH model in explicit state space form are constructed as described before on a regular simplicial triangulation of Ω. The simulations results, which are shown below, are performed on a grid with step size $h_x = h_y = \frac{1}{48}$.

The simulation snapshots both in Fig. 4.16 and Fig. 4.17 show the expected wave propagation, which produces an interference pattern on the square part $(0, 1) \times (0, 1)$ of Ω. Due to the rather coarse grid, the differences between the two considered parametrizations become evident. For parameter set #3, a remarkable distortion to the left of the wave front occurs. For parameter set #4, the distribution of e^p in x direction along the narrow strips is less uniform, with a relatively large deviation on the right hand boundary.

A remedy for the two described phenomena can be "symmetrization" of the mesh in the sense that the square cells of the grid are divided by diagonal edges with alternating orientation (lower left to upper right vs. lower right

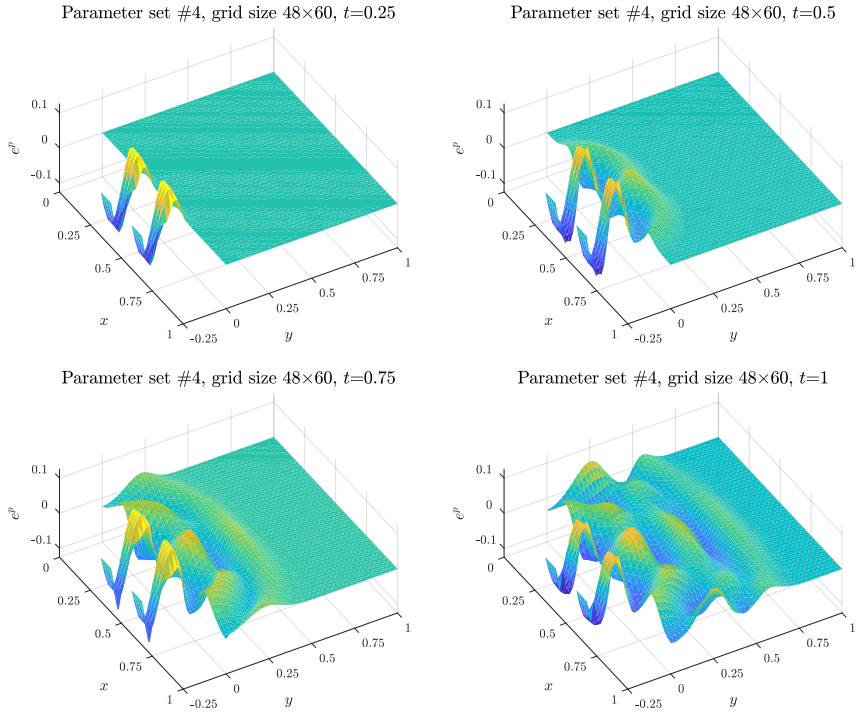

Figure 4.17: Snapshots of the double slit experiment with parameter set #4.

to upper left). This modification of the grid topology requires also a modified construction of power-preserving mappings, which is an issue of ongoing implementation work.

4.8 Conclusions

We introduced the weak form of the Stokes-Dirac structure with a segmented boundary, on which the causality of the port variables (the assignment as system in- or outputs) alternates. This Stokes-Dirac structure is the underlying geometric structure to represent power continuity in a port-Hamiltonian distributed parameter system. On the example of a system of two conservation laws with canonical interdomain coupling, we described the mixed Galerkin discretization of the Stokes-Dirac structure in a general way. To obtain finite-dimensional approximate models in PH form with the prescribed boundary inputs – as a basis for the interconnection of multi-physics models, control design and simulation – we proposed power-preserving mappings on the space of discrete effort and flow variables. These maps allow to define non-degenerate duality pairings, leading to finite-dimensional approximate Dirac structures on the minimal discrete bond space. The Dirac structures admit several repre-

sentations, one of them being an explicit input-output-representation. Port-Hamiltonian state space models are obtained, if dynamics is added and the constitutive equations are approximated consistently. On the example of Whitney finite elements we demonstrated the discretization procedure in 1D and 2D and gave interpretations of the resulting discretization schemes.

The proposed method is, to the best of our knowledge, the first method which allows for a structure-preserving discretization of PH distributed parameter systems in more than one spatial dimension with a systematic treatment of different boundary inputs and the possibility to tune the discretized models between centered schemes and upwinding. The proposed family of approximation Dirac structures avoids a direct feedthrough in the state space model of the discretized wave equation and the over-estimation of higher frequencies in the approximate spectrum, which is the case for the method presented in [71], where the efforts instead of the flow degrees of freedom are mapped. The weak form of the Stokes-Dirac structure is the key feature that allows to include additional effects such as dissipation or diffusion or, more generally, to tackle the discretization of PH systems with general and higher order interconnection operators and distributed inputs.

An important difference of our work to related works like [174], [56], [82], where either dual grids are used *a priori* or at least one conservation law contains the Hodge star or the co-differential, is that our initial discretization is based on a *metric-independent* formulation of the conservation laws. We approximate all differential forms in the same conforming subspaces depending on their degree (i. e. on the same mesh in FE), which has the advantage that boundary variables are defined directly *on* $\partial\Omega$, without having to cope with an eventual grid shift. To obtain an explicit state space model, however, we need – *no free lunch* – the power-preserving mappings. These, in turn, give us degrees of freedom to tune the resulting numerical method.

Current and future work concerns the application of the method to the PH representations of systems including heat and mass diffusion phenomena, which share similar Stokes-Dirac structures, as well as coupled heat and mass transport phenomena in non-homogeneous media such as catalytic foams, see Section 3.5. Moreover, we want to analyze the approach when applied to PH systems with non-canonical system operators (containing e. g. higher order derivatives). In this context, we are interested in the reasonable choice of design parameters in order to adapt the discretization scheme to the physical nature of the system (e. g. to account for the ratio between convection and diffusion). This aspect is closely related to the analysis of system-theoretic properties of the discretized models in view of control design. Further important issues are the implementation of the approach in existing finite element tools like FEniCS [2] and the use of approximation spaces with higher degree [163], [7]. We intend to include the discretization of the nonlinear constitutive relations for the 2D shallow water equations in our *open* models and clarify the links with recent work on geometric mixed finite elements like [41], [42], where in- and outputs are not explicitly taken into account, and upwinding in differential forms as presented in [39].

Chapter 5

Structure-Preserving Time Discretization

We introduce a new definition of discrete-time port-Hamiltonian (PH) systems, which results from structure-preserving time discretization of explicit finite-dimensional PH systems[1]. We discretize the underlying continuous-time Dirac structure with the *collocation method* and add discrete-time dynamics by the use of *symplectic numerical integration* schemes. The conservation of a discrete-time energy balance – expressed in terms of the discrete-time Dirac structure – extends the notion of symplecticity of geometric integration schemes to (open) control systems. The quadrature formulas, which are associated with the polynomial approximations of the power variables, allow for quantitative statements on the approximation error of the solution, the supplied and the stored energy. We show that among collocation methods only Gauss-Legendre collocation, which leads to implicit multi-stage Runge-Kutta schemes with maximum order, guarantees an *exact* discrete energy balance as defined in [36], Def. III.2, if applied to linear PH systems. Our definition includes discretization schemes, which yield a non-exact but consistent discrete energy balance. An example are the Lobatto IIIA/IIIB pairs for partitioned systems. The energy errors are then *consistent with*, i. e. they have the same order as the chosen integration scheme. The statements on the numerical energy errors are illustrated by elementary numerical experiments.

The chapter is organized as follows. In Section 5.1, we recall the considered class of finite-dimensional PH systems with their underlying Dirac structure. Section 5.2 contains as main results the definitions of discrete-time Dirac structures and PH systems based on the collocation method. In Section 5.3, we consider Gauss-Legendre methods and Lobatto IIIA/IIIB pairs and discuss the orders of the energy approximations. Section 5.4 illustrates the statements on the elementary example of a linear undamped/damped oscillator. In the concluding Section 5.5, we summarize the chapter, and we point out perspectives

[1] This chapter corresponds to the slightly edited initial version of [102].

for future work based on the presented results.

The chapter further develops earlier results on the symplectic time integration of PH systems using Gauss-Legendre collocation [101]. The main novelties are the precise consideration of the different energy approximations, the application of the ideas to s-stage Lobatto pairs for partitioned systems, the analysis and order proofs for the energy errors, and the extended section on numerical experiments.

5.1 Lossless Port-Hamiltonian Systems

We consider the class of *lossless* finite-dimensional PH systems in an *explicit* input-state-output representation as introduced in Subsection 2.1.2,

$$\dot{\mathbf{x}}(t) = \mathbf{J}(\mathbf{x}(t))\nabla H(\mathbf{x}(t)) + \mathbf{G}(\mathbf{x}(t))\mathbf{u}(t) \tag{5.1a}$$

$$\mathbf{y}(t) = \mathbf{G}^T(\mathbf{x}(t))\nabla H(\mathbf{x}(t)), \tag{5.1b}$$

with state vector $\mathbf{x} \in \mathbb{R}^n$, collocated in- and output vectors $\mathbf{u}, \mathbf{y} \in \mathbb{R}^m$. The Hamiltonian $H : \mathbb{R}^n \to \mathbb{R}$ is bounded from below with a strict minimum in $\mathbf{x}^*$, which is the equilibrium state for $\mathbf{u} \equiv \mathbf{0}$. By skew-symmetry of the interconnection matrix $\mathbf{J} = -\mathbf{J}^T$ and the definition of the collocated output, the differential energy balance

$$\dot{H}(\mathbf{x}(t)) = \mathbf{y}^T(t)\mathbf{u}(t) \tag{5.2}$$

holds for all t, or in integral form,

$$H(\mathbf{x}(t_2)) - H(\mathbf{x}(t_1)) = \int_{t_1}^{t_2} \mathbf{y}^T(s)\mathbf{u}(s)\,ds, \quad \forall t_1 \le t_2, \tag{5.3}$$

which shows passivity (see e. g. [191]) of the state representation (5.1). The energy balance is a *structural* or *geometric* property, i. e. it holds independently of $H(\mathbf{x})$. *Flow* and *effort* vectors are defined as

$$\mathbf{f}(t) := -\dot{\mathbf{x}}(t), \qquad \mathbf{e}(t) := \nabla H(\mathbf{x}(t)). \tag{5.4}$$

Because of $\dot{H} = (\nabla H)^T\dot{\mathbf{x}} = -\mathbf{e}^T\mathbf{f}$, they represent *power-conjugated*, *dual* variables. The differential energy balance (5.2) can be written as the power balance equation on the *bond space* $\mathcal{F} \times \mathcal{E}$, with $\mathcal{F} = \mathbb{R}^n \times \mathbb{R}^m \ni (\mathbf{f}, \mathbf{y})$ and $\mathcal{E} = \mathbb{R}^n \times \mathbb{R}^m \ni (\mathbf{e}, \mathbf{u})$:

$$\mathbf{e}^T(t)\mathbf{f}(t) + \mathbf{y}^T(t)\mathbf{u}(t) = 0. \tag{5.5}$$

In integral form, we obtain a *structural* energy balance over every interval $[t_1, t_2]$:

$$\int_{t_1}^{t_2} \mathbf{e}^T(s)\mathbf{f}(s) + \mathbf{y}^T(s)\mathbf{u}(s)\,ds = 0. \tag{5.6}$$

By

$$-\mathbf{f}(t) = \mathbf{J}(\mathbf{x}(t))\mathbf{e}(t) + \mathbf{G}(\mathbf{x}(t))\mathbf{u}(t) \tag{5.7a}$$

$$\mathbf{y}(t) = \mathbf{G}^T(\mathbf{x}(t))\mathbf{e}(t), \tag{5.7b}$$

the bond variables are constrained to a subspace (i. e. the graph of the skew-symmetric map defined in (5.7)), on which in particular (5.5) holds. This subspace is called a *Dirac structure*. For more details on Dirac structures and PH systems, see e. g. [191], Chapter 6.

5.2 Discrete-Time PH Systems Based on Collocation

We define the class of *discrete-time PH systems*, which arise from a *discrete-time Dirac structure*. The latter is obtained by applying the collocation method to the class of PH systems (5.1) and by defining in an appropriate manner *discrete flow and effort vectors* for every sampling interval. Special attention is paid on the discretization of the energy balance (5.3). For a *consistent* discrete-time approximation of the PH system (5.1), both the *supplied energy* (right hand side of (5.3)) and the *stored energy* (left hand side of (5.3)) must be approximated with the same order.

5.2.1 Collocation Method

We consider equidistant sampling intervals $I^k = [t_0^k, t_{s+1}^k] = [(k-1)h, kh]$, $k \in \mathbb{N}$ for the time t with $t_{s+1}^k = t_0^k + h$, see Fig. 5.1. With $t = ((k-1)+\tau)h$, the sampling intervals are parametrized by the normalized time $\tau \in [0,1]$. The polynomial approximations of the system variables will be denoted with a tilde. As described in Section II.1.2 of [76], the numerical approximation of the solution $\mathbf{x}(t)$ of (5.1) is given by the vector $\tilde{\mathbf{x}}(t) \in \mathbb{R}^n$ of *collocation polynomials* of degree s. Assume first the initial value $\mathbf{x}_0^k := \tilde{\mathbf{x}}(t_0^k) = \mathbf{x}(t_0^k)$ to be known. The continuous numerical solution $\tilde{\mathbf{x}}(t)$ is then the vector of polynomials whose time derivative matches the right hand side of (5.1a) in the collocation points $t_i^k := t_0^k + c_i h$, $i = 1, \dots s$, with $0 \le c_i \le 1$:

$$\begin{aligned} \dot{\tilde{\mathbf{x}}}(t_i^k) &= -\mathbf{f}_i^k, \\ -\mathbf{f}_i^k &= (\mathbf{J}(\mathbf{x})\nabla H(\mathbf{x}))|_{\mathbf{x}=\tilde{\mathbf{x}}(t_i^k)} + \mathbf{G}(\tilde{\mathbf{x}}(t_i^k))\mathbf{u}(t_i^k). \end{aligned} \tag{5.8}$$

Notation: Arguments in *latin* letters (t or s under the integral) refer to time functions evaluated on I^k. *Greek* letters (τ or σ) refer to the same function, mapped to the normalized interval $[0,1]$.

5.2.2 Approximation of Flow and State Variables

Based on $\mathbf{f}_i^k \in \mathbb{R}^n$, $i = 1, \dots, s$, according to (5.8), the interpolation formula

$$\dot{\tilde{\mathbf{x}}}(t_0^k + \tau h) =: -\tilde{\mathbf{f}}(t_0^k + \tau h) = -\sum_{i=1}^{s} \mathbf{f}_i^k \ell_i(\tau), \tag{5.9}$$

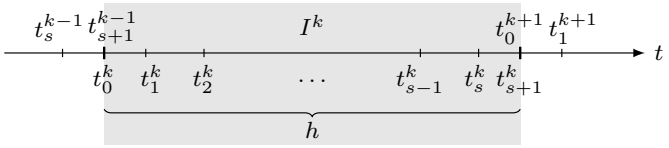

Figure 5.1: Sampling interval I^k with interior collocation points $t_i^k = t_0^k + c_i h$, $i = 1, \ldots, s$.

with ℓ_i the i-th Lagrange interpolation polynomial

$$\ell_i(\tau) = \prod_{\substack{j=1 \\ j \neq i}}^{s} \frac{\tau - c_j}{c_i - c_j}, \qquad \tau \in [0,1], \tag{5.10}$$

gives a polynomial approximation of $\dot{\mathbf{x}}(t)$ on I^k. The flow coordinates are merged in the *discrete-time flow vector*

$$\mathbf{f}^k := \left[(\mathbf{f}_1^k)^T \quad \ldots \quad (\mathbf{f}_s^k)^T \right]^T \in \mathbb{R}^{sn}, \tag{5.11}$$

based on which the numerical solution $\tilde{\mathbf{x}}(t^k + \tau h)$, $\tau \in [0,1]$ is obtained by integration of (5.9):

$$\tilde{\mathbf{x}}(t_0^k + \tau h) = \tilde{\mathbf{x}}(t_0^k) - h \sum_{j=1}^{s} \Big(\mathbf{f}_j^k \int_0^\tau \ell_j(\sigma)\, d\sigma \Big). \tag{5.12}$$

The values $\mathbf{x}_i^k := \tilde{\mathbf{x}}(t_i^k)$ of the numerical solution inside and at the end of the interval I^k are then computed as

$$\mathbf{x}_i^k = \mathbf{x}_0^k - h \sum_{j=1}^{s} a_{ij} \mathbf{f}_j^k, \quad i = 1, \ldots, s, \tag{5.13}$$

$$\mathbf{x}_{s+1}^k = \mathbf{x}_0^k - h \sum_{j=1}^{s} b_j \mathbf{f}_j^k, \tag{5.14}$$

with[2] $(i, j = 1, \ldots, s)$

$$a_{ij} = \int_0^{c_i} \ell_j(\sigma)\, d\sigma, \qquad b_j = \int_0^1 \ell_j(\sigma)\, d\sigma. \tag{5.15}$$

In *continuous* collocation methods, the numerical solution at the start $t_0^{k+1} = t_{s+1}^k$ of the subsequent interval is initialized by $\mathbf{x}_0^{k+1} = \mathbf{x}_{s+1}^k$.

[2] These values are, together with c_i, the coefficients of the Butcher table for the Runge-Kutta (RK) interpretation of the collocation method, see [76], Theorem II.1.4.

5.2.3 Effort Approximation and Discrete-Time Structure Equation

The definition of the discrete flow coordinates $\mathbf{f}_1^k, \ldots, \mathbf{f}_s^k$ in (5.8) requires to evaluate the effort vector $\nabla H(\mathbf{x}(t))$, the input $\mathbf{u}(t)$ and the interconnection and input matrices $\mathbf{J}(\mathbf{x}(t))$, $\mathbf{G}(\mathbf{x}(t))$ *in the flow collocation points* c_i,

$$\mathbf{e}_i^k := \nabla H(\mathbf{x})|_{\mathbf{x}_i^k}, \qquad \mathbf{u}_i^k := \mathbf{u}(t_0^k + c_i h), \tag{5.16}$$

and

$$\mathbf{J}_i^k := \mathbf{J}(\mathbf{x}_i^k), \qquad \mathbf{G}_i^k := \mathbf{G}(\mathbf{x}_i^k) \tag{5.17}$$

for $i = 1, \ldots, s$. The discrete-time counterpart of (5.7a) is then

$$-\mathbf{f}_i^k = \mathbf{J}_i^k \mathbf{e}_i^k + \mathbf{G}_i^k \mathbf{u}_i^k, \tag{5.18}$$

with $\mathbf{J}_i^k = -(\mathbf{J}_i^k)^T$. With $\mathbf{e}_i^k \in \mathbb{R}^n$ and $\mathbf{u}_i^k \in \mathbb{R}^m$ the *discrete effort* and *discrete input coordinates* according to (5.16), the polynomial approximations of the effort and the input vector are

$$\tilde{\mathbf{e}}(t_0^k + \tau h) = \sum_{i=1}^{s} \mathbf{e}_i^k \ell_i(\tau), \qquad \tilde{\mathbf{u}}(t_0^k + \tau h) = \sum_{i=1}^{s} \mathbf{u}_i^k \ell_i(\tau). \tag{5.19}$$

In accordance with the approximate flows, we define the *discrete-time effort vector* and the *discrete-time input vector*

$$\mathbf{e}^k = \begin{bmatrix} (\mathbf{e}_1^k)^T & \ldots & (\mathbf{e}_s^k)^T \end{bmatrix}^T \in \mathbb{R}^{sn}, \tag{5.20}$$

$$\mathbf{u}^k = \begin{bmatrix} (\mathbf{u}_1^k)^T & \ldots & (\mathbf{u}_s^k)^T \end{bmatrix}^T \in \mathbb{R}^{sm}. \tag{5.21}$$

Defining the block-diagonal matrices

$$\begin{aligned} \mathbf{J}^k = -(\mathbf{J}^k)^T &= \operatorname{blockdiag}(\mathbf{J}_1^k, \ldots, \mathbf{J}_s^k), \\ \mathbf{G}^k &= \operatorname{blockdiag}(\mathbf{G}_1^k, \ldots, \mathbf{G}_s^k), \end{aligned} \tag{5.22}$$

the *structure equation* (5.18) on the sampling interval I^k can be rewritten as

$$-\mathbf{f}^k = \mathbf{J}^k \mathbf{e}^k + \mathbf{G}^k \mathbf{u}^k. \tag{5.23}$$

Remark 5.1. Defining the discrete effort and input coordinates based on *different* collocation points $d_1, \ldots, d_r$ is also conceivable. In this case, the terms of the right hand side of (5.18) must be replaced by interpolations between the effort collocation points according to (5.19). In this paper, we restrict ourselves to *identical* collocation points for the flow and effort variables, and to the explicit representation of the resulting Dirac structure and PH system.

5.2.4 Discrete-Time Supplied Energy

Since the instantaneous (local) power balance results trivially from the equations of the Dirac structure, we seek to express a discrete-time counterpart of the integral energy balance equation (5.3) on the time interval I^k. To this end, we integrate the polynomial approximation of instantaneous power $-\tilde{\mathbf{e}}^T(t_0^k+\tau h)\tilde{\mathbf{f}}(t_0^k+\tau h)$ over the normalized time interval $[0,1]$, and obtain an *approximation of supplied energy* on the sampling interval I^k

$$\Delta\tilde{H}^k := -h\int_0^1 \tilde{\mathbf{e}}^T(t_0^k+\sigma h)\tilde{\mathbf{f}}(t_0^k+\sigma h)\,d\sigma. \tag{5.24}$$

Substituting the definitions (5.9) and (5.19) of the polynomial flow and effort approximations, we obtain the bilinear form

$$\Delta\tilde{H}^k = -h(\mathbf{e}^k)^T\mathbf{M}\mathbf{f}^k \tag{5.25}$$

with the symmetric matrix $\mathbf{M}=\mathbf{M}^T\in\mathbb{R}^{sn\times sn}$,

$$\mathbf{M}=\begin{bmatrix} m_{11} & \dots & m_{1s}\\ \vdots & \ddots & \vdots\\ m_{s1} & \dots & m_{ss}\end{bmatrix}\otimes\mathbf{I}_n,\qquad m_{ij}=\int_0^1 \ell_i(\sigma)\ell_j(\sigma)d\sigma, \tag{5.26}$$

where $m_{ij}=m_{ji}$ and $\otimes$ denotes the Kronecker product.

Remark 5.2. We can understand the term $\mathbf{M}\mathbf{e}^k$ as a generalization of the *discrete gradient* and $-h\mathbf{f}^k$ as a vector generalizing the increment of the numerical solution in the integration step.

5.2.5 Discrete-Time Dirac Structure

We provide conditions under which the polynomial approximation of the power variables leads to the definition of a discrete-time Dirac structure. Substituting the relation (5.23) in the right hand side of the discrete energy balance (5.25), we obtain

$$-h(\mathbf{e}^k)^T\mathbf{M}\mathbf{f}^k = h(\mathbf{e}^k)^T\mathbf{M}\mathbf{J}^k\mathbf{e}^k + h(\mathbf{e}^k)^T\mathbf{M}\mathbf{G}^k\mathbf{u}^k. \tag{5.27}$$

At this stage, we want to recover a discrete-time equivalent of the structural power balance (5.5). To this end, the first term on the right hand side must vanish: $h(\mathbf{e}^k)^T\mathbf{M}\mathbf{J}^k\mathbf{e}^k \stackrel{!}{=} 0$ for all $\mathbf{e}^k\in\mathbb{R}^{sn}$, or written element-wise (recall $m_{ij}=m_{ji}$),

$$h\begin{bmatrix}(\mathbf{e}_1^k)^T & \dots & (\mathbf{e}_s^k)^T\end{bmatrix}\begin{bmatrix} m_{11}\mathbf{I}_n & \dots & m_{1s}\mathbf{I}_n\\ \vdots & \ddots & \vdots\\ m_{s1}\mathbf{I}_n & \dots & m_{ss}\mathbf{I}_n\end{bmatrix}\begin{bmatrix}\mathbf{J}_1^k & & \\ & \ddots & \\ & & \mathbf{J}_s^k\end{bmatrix}\begin{bmatrix}\mathbf{e}_1^k\\ \vdots\\ \mathbf{e}_s^k\end{bmatrix}$$
$$= h\sum_{i=1}^{s}\sum_{j=1}^{s}(\mathbf{e}_i^k)^T m_{ij}\mathbf{J}_j^k\mathbf{e}_j^k \stackrel{!}{=} 0. \tag{5.28}$$

By skew-symmetry of $\mathbf{J}_j^k$, we have $(\mathbf{e}_j^k)^T m_{jj} \mathbf{J}_j^k \mathbf{e}_j^k = 0$ for all $j = 1, \ldots, s$. The remaining elements of the sum cancel to zero, if $(\mathbf{e}_i^k)^T m_{ij} \mathbf{J}_j^k \mathbf{e}_j^k = -(\mathbf{e}_j^k)^T m_{ji} \mathbf{J}_i^k \mathbf{e}_i^k$ holds for all $i \neq j$. With the equality $(\mathbf{e}_i^k)^T m_{ij} \mathbf{J}_j^k \mathbf{e}_j^k = -((\mathbf{e}_i^k)^T m_{ij} (\mathbf{J}_j^k)^T \mathbf{e}_j^k)^T = -(\mathbf{e}_j^k)^T m_{ji} \mathbf{J}_j^k \mathbf{e}_i^k$, the requirement translates to

$$(\mathbf{e}_j^k)^T m_{ji} \mathbf{J}_i^k \mathbf{e}_i^k = (\mathbf{e}_j^k)^T m_{ji} \mathbf{J}_j^k \mathbf{e}_i^k \qquad \forall i \neq j, \tag{5.29}$$

which is true if either one of the following conditions holds.

(C1) $m_{ij} = 0$ for all $i \neq j$,

(C2) $\mathbf{J}_i^k = \mathbf{J}_j^k = const.$ for all $i, j = 1, \ldots, s$.

While (C1) is an orthogonality condition on the choice of the approximation basis in the *discretization method*, the constant interconnection structure according to (C2) is a *system property*. In both cases (C1) or (C2), the discrete energy balance (5.27) boils down to

$$h(\mathbf{e}^k)^T \mathbf{M} \mathbf{f}^k + h(\mathbf{e}^k)^T \mathbf{M} \mathbf{G}^k \mathbf{u}^k = 0. \tag{5.30}$$

The definition of a *discrete-time output vector*

$$\mathbf{y}^k := (\mathbf{G}^k)^T \mathbf{M} \mathbf{e}^k \tag{5.31}$$

yields (using $\mathbf{M} = \mathbf{M}^T$)

$$h(\mathbf{M}\mathbf{e}^k)^T \mathbf{f}^k + h(\mathbf{y}^k)^T \mathbf{u}^k = 0, \tag{5.32}$$

which represents a *structural balance equation* for the supplied energy in terms of the discrete-time conjugate port variables in the nodes of the sampling interval I^k. We are ready to define the discrete-time Dirac structure, which is based on the polynomial approximation of the power variables.

Theorem 5.1 (Discrete-time Dirac structure)**.** Given the s collocation points $0 \leq c_i \leq 1$, $i = 1, \ldots, s$. The system of equations (5.23), (5.31), i. e.

$$\begin{aligned} -\mathbf{f}^k &= \mathbf{J}^k \mathbf{e}^k + \mathbf{G}^k \mathbf{u}^k \\ \mathbf{y}^k &= (\mathbf{G}^k)^T \mathbf{M} \mathbf{e}^k, \end{aligned} \tag{5.33}$$

with discrete flow, effort and input vectors $\mathbf{f}^k$, $\mathbf{e}^k$, $\mathbf{u}^k$ according to (5.11), (5.20), (5.21), the block matrices $\mathbf{J}^k = -(\mathbf{J}^k)^T$, $\mathbf{G}^k$ according to (5.22), and the symmetric block matrix $\mathbf{M} = \mathbf{M}^T$ according to (5.26), represents a *discrete-time Dirac structure* on the time interval $I^k = [(k-1)h, kh]$, which approximates the continuous-time Dirac structure according to (5.7), if either condition (C1) or (C2) is satisfied.

Proof. Define $\hat{\mathbf{f}}^k := \mathbf{M}\mathbf{f}^k$. We can write (5.33) as

$$\begin{bmatrix} \hat{\mathbf{f}}^k \\ \mathbf{y}^k \end{bmatrix} + \begin{bmatrix} \mathbf{M}\mathbf{J}^k & \mathbf{M}\mathbf{G}^k \\ -(\mathbf{G}^k)^T\mathbf{M} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}^k \\ \mathbf{u}^k \end{bmatrix} = \mathbf{0}, \tag{5.34}$$

which is the *kernel* representation of a finite-dimensional Dirac structure according to Theorem 2.2 if the matrix in the second term is skew-symmetric. This is the case if $\mathbf{M}\mathbf{J}^k$ is skew-symmetric, for instance when conditions (C1) or (C2) hold. On the Dirac structure, the balance equation

$$(\mathbf{e}^k)^T\hat{\mathbf{f}}^k + (\mathbf{y}^k)^T\mathbf{u}^k = 0 \tag{5.35}$$

holds. Substituting the definitions of $\hat{\mathbf{f}}^k$ and $\mathbf{y}^k$, and multiplying with h, we obtain (5.30), which is a quadrature formula for the structural energy balance (5.6). □

5.2.6 Discrete-Time Port-Hamiltonian System

The discrete-time Dirac structure is now complemented by discrete-time dynamics and constitutive equations.

Definition 5.1 (Discrete-time PH system)**.** Equations (5.33), together with the *s-stage discrete dynamics*

$$\mathbf{x}_0^k = \mathbf{x}_{s+1}^{k-1}, \tag{5.36a}$$

$$\mathbf{x}_i^k = \mathbf{x}_0^k - h\sum_{j=1}^{s} a_{ij}\mathbf{f}_j^k, \qquad i = 1, \dots, s, \tag{5.36b}$$

$$\mathbf{x}_{s+1}^k = \mathbf{x}_0^k - h\sum_{j=1}^{s} b_j\mathbf{f}_j^k, \tag{5.36c}$$

with Runge-Kutta coefficients a_{ij} and b_j according to (5.15) and the *discrete constitutive equations*

$$\mathbf{e}_i^k = \nabla H(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_i^k}, \qquad i = 1, \dots, s, \tag{5.37}$$

define a discrete-time dynamical system. Using (5.25) and (5.32), the approximation of *supplied energy* on the sampling interval $I^k = [(k-1)h, kh]$ is given by

$$\Delta\tilde{H}^k = h(\mathbf{y}^k)^T\mathbf{u}^k. \tag{5.38}$$

The so-defined dynamical system is called a *discrete-time PH system* if the error between $\Delta\tilde{H}^k$ and the *increment of stored energy*

$$\Delta\bar{H}^k = H(\mathbf{x}_{s+1}^k) - H(\mathbf{x}_0^k) \tag{5.39}$$

has the order of the integration scheme (5.36).

Remark 5.3. Equations (5.36b) and (5.36c) can be written in more compact form using the Kronecker product

$$\mathbf{x}_i^k = \mathbf{x}_0^k - h(\mathbf{a}_i^T \otimes \mathbf{I}_n)\mathbf{f}^k, \qquad i = 1, \ldots, s, \tag{5.40a}$$

$$\mathbf{x}_{s+1}^k = \mathbf{x}_0^k - h(\mathbf{b}^T \otimes \mathbf{I}_n)\mathbf{f}^k, \tag{5.40b}$$

with $\mathbf{a}_i^T = \begin{bmatrix} a_{i1} & \ldots & a_{is} \end{bmatrix}$, $\mathbf{b}^T = \begin{bmatrix} b_1 & \ldots & b_s \end{bmatrix}$.

Remark 5.4. The *structure* of the discrete-time PH models is independent of the sampling time h, which, however, determines the approximation quality.

5.2.7 Discrete Energy Balance

Equation (5.38) represents an approximation of the *supplied energy* flow through the port $(\mathbf{u}(t), \mathbf{y}(t))$ based on the polynomial approximations of flows and efforts. On the other hand, (5.39) expresses the increment of *stored energy* based on a numerical integration scheme (5.36), as opposed to the exact increment

$$\Delta H^k = H(\mathbf{x}(t_0^k + h)) - H(\mathbf{x}(t_0^k)). \tag{5.41}$$

Definition 5.2 (Consistent energy balance)**.** If under a given numerical integration scheme of order p, the increment of stored energy satisfies

$$\Delta \bar{H}^k = h(\mathbf{y}^k)^T \mathbf{u}^k + o(h^p), \tag{5.42}$$

we call (5.42) a *discrete energy balance*, which is *consistent* with the discretization scheme. If

$$\Delta \bar{H}^k = h(\mathbf{y}^k)^T \mathbf{u}^k, \tag{5.43}$$

we call the energy balance *exact*.

Remark 5.5. In Definition III.2 of [36], the latter case is simply called "discrete energy balance". As (5.43) only holds under additional conditions, like constant structure matrix and quadratic energy, and for example under the implicit midpoint rule (see [4], Section III.B and [36], Section III.E), we add "exact" to distinguish this particular case of a discrete energy balance.

To show the consistency of a discrete energy balance, we will show that the local approximation errors of both $\Delta \bar{H}^k$ and $\Delta \tilde{H}^k$, compared with the exact increment ΔH^k with $\mathbf{x}_0^k = \mathbf{x}(t_0^k)$, have order $o(h^p)$. To perform this analysis, we restrict ourselves to quadratic energies of the form

$$H(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x}, \qquad \mathbf{Q} = \mathbf{Q}^T > 0. \tag{5.44}$$

Theorem 5.2 (Local error of stored energy)**.** For a linear PH system with quadratic energy (5.44), the local energy error

$$\Delta\bar{H}^k - \Delta H^k, \qquad \mathbf{x}_0^k = \mathbf{x}(t_0^k), \tag{5.45}$$

is consistent with the numerical integration scheme, i. e. it has order $o(h^p)$.

Proof. For an integration scheme of order p, the *local* or consistency error (set $\mathbf{x}_0^k = \tilde{\mathbf{x}}(t_0^k) = \mathbf{x}(t_0^k)$) has order $o(h^p)$: $\|\tilde{\mathbf{x}}(t_0^k + h) - \mathbf{x}(t_0^k + h)\| \leq C_1 h^{p+1}$, $C_1 > 0$. By the equivalence of norms, this holds accordingly for the energy norm $\|\mathbf{x}\|_Q := \sqrt{\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}}$, with a different constant $C_2 > 0$: $\|\tilde{\mathbf{x}}(t_0^k + h) - \mathbf{x}(t_0^k + h)\|_Q \leq C_2 h^{p+1}$. For the error in the stored energy, the following estimate can be given, where the triangle inequality is used between third and fourth line ($t_{s+1}^k = t_0^k + h$, $\mathbf{x}_{s+1}^k = \tilde{\mathbf{x}}(t_{s+1}^k)$):

$$\begin{aligned}
\left|\Delta\bar{H}^k - \Delta H^k\right| &= \left|\|\mathbf{x}_{s+1}^k\|_Q^2 - \|\mathbf{x}(t_{s+1}^k)\|_Q^2\right| \\
&= \left|\left(\|\mathbf{x}_{s+1}^k\|_Q + \|\mathbf{x}(t_{s+1}^k)\|_Q\right) \cdot \left(\|\mathbf{x}_{s+1}^k\|_Q - \|\mathbf{x}(t_{s+1}^k)\|_Q\right)\right| \\
&\leq 2\max\left(\|\mathbf{x}_{s+1}^k\|_Q, \|\mathbf{x}(t_{s+1}^k)\|_Q\right) \cdot \left|\|\mathbf{x}_{s+1}^k\|_Q - \|\mathbf{x}(t_{s+1}^k)\|_Q\right| \\
&\leq 2\max\left(\|\mathbf{x}_{s+1}^k\|_Q, \|\mathbf{x}(t_{s+1}^k)\|_Q\right) \cdot \left\|\mathbf{x}_{s+1}^k - \mathbf{x}(t_{s+1}^k)\right\|_Q \\
&\leq C_3 h^{p+1}
\end{aligned}$$

with $C_3 = 2\max\left(\|\mathbf{x}_{s+1}^k\|_Q, \|\mathbf{x}(t_{s+1}^k)\|_Q\right) C_2$. □

In the following section, we keep the assumption of quadratic energies and discuss Gauss-Legendre collocation and the Lobatto IIIA/IIIB pairs for partitioned systems as examples for symplectic integration schemes. In the former case, we will prove that $\Delta\bar{H}^k = \Delta\tilde{H}^k$ holds, while in the latter case, we will show that $\Delta\tilde{H}^k$ approximates the energy increment ΔH^k with an error of identical order as $\Delta\bar{H}^k$, and consequently the (consistent) discrete energy balance (5.42) holds.

5.3 Examples and Analysis of Energy Errors

The degrees of freedom to define a discrete-time PH system according to Definition 5.1 are the set of collocation points $\{c_1, \ldots, c_s\}$ and the concrete coefficients a_{ij} and weights b_j of the RK integration scheme (5.36). In this Section, we analyze the error of $\Delta\tilde{H}^k$ and its relation to $\Delta\bar{H}^k$ in the light of Definition 5.2 for both Gauss-Legendre collocation and the Lobatto IIIA/IIIB pairs.

5.3.1 Gauss-Legendre Collocation

To define a discrete-time Dirac structure for cases with non-constant interconnection matrices, condition (C1) must be satisfied by the choice of collocation points. To have $m_{ij} = m_{ji} = 0$ for $i \neq j$, with m_{ij} defined in (5.26), the

interpolation polynomials $\{\ell_1(\tau), \ldots, \ell_s(\tau)\}$ must form a system of *orthogonal* functions. This is the case, if we take the collocation points $c_1, \ldots, c_s$, as the zeros of the shifted Legendre polynomials[3]

$$\frac{d^s}{d\tau^s}\,(\tau^s(\tau-1)^s) \tag{5.46}$$

in normalized time $\tau \in [0,1]$. With the resulting interpolation polynomials, the diagonal elements of $\mathbf{M}$,

$$m_{ii} = \int_0^1 \ell_i^2(\sigma)\,d\sigma = \int_0^1 \ell_i(\sigma)\,d\sigma = b_i, \tag{5.47}$$

equal the weights b_i in the Gauss-Legendre quadrature formula

$$\int_0^1 f(\sigma)\,d\sigma \approx \sum_{i=1}^{s} b_i f(c_i). \tag{5.48}$$

This quadrature formula is *exact* for polynomials $f(\tau)$ on $[0,1]$ up to order $2s-1$. With $2s$ the order of the quadrature formula, the approximation error of the integral

$$\int_{t_0}^{t_0+h} f(s)\,ds = h\int_0^1 f(t_0+h\sigma)\,d\sigma \approx h\sum_{i=1}^{s} b_i f(t_0+hc_i) \tag{5.49}$$

is of order $\mathcal{O}(h^{2s+1})$. The coefficients a_{ij} of the unique implicit Runge-Kutta (RK) method[4] of order $p=2s$ can be computed as given in (5.15).

We now determine the conditions on the parameters of the integration scheme under which $\Delta\bar{H}^k = \Delta\tilde{H}^k$. Substituting (5.36c) in (5.39) for a quadratic energy, we have

$$\Delta\bar{H}^k = -h(\mathbf{x}_0^k)^T\mathbf{Q}\sum_{j=1}^{s} b_j\mathbf{f}_j^k + \frac{1}{2}h^2(\sum_{j=1}^{s} b_j\mathbf{f}_j^k)^T\mathbf{Q}\sum_{j=1}^{s} b_j\mathbf{f}_j^k. \tag{5.50}$$

On the other hand, with $b_j = m_{jj}$, $\mathbf{e}_j^k = \mathbf{Q}\mathbf{x}_j^k$ and (5.36c), Equation (5.25) becomes

$$\begin{aligned}\Delta\tilde{H}^k &= -h\sum_{j=1}^{s}(\mathbf{e}_j^k)^T m_{jj}\mathbf{f}_j^k \\ &= -h\sum_{j=1}^{s}(\mathbf{x}_0^k - h\sum_{l=1}^{s} a_{jl}\mathbf{f}_j^k)^T\mathbf{Q}b_j\mathbf{f}_j^k \\ &= -h(\mathbf{x}_0^k)^T\mathbf{Q}\sum_{j=1}^{s} b_j\mathbf{f}_j^k + h^2\sum_{j=1}^{s}\Big((\sum_{l=1}^{s} a_{jl}\mathbf{f}_l^k)^T\mathbf{Q}b_j\mathbf{f}_j^k\Big).\end{aligned} \tag{5.51}$$

[3] See [76], Section II.1.3.

[4] See the Butcher tables in Appendix A.2.

The first terms in (5.50) and (5.51) are identical. By matching the coefficients in front of $h^2(\mathbf{f}_l^k)^T\mathbf{Q}\mathbf{f}_j^k$ in the second term, one obtains the conditions

$$a_{ii} = \frac{1}{2}b_i, \qquad a_{ij}b_i + a_{ji}b_j = b_i b_j \tag{5.52}$$

for $i, j = 1, \ldots, s$, under which $\Delta\bar{H}^k$ and $\Delta\tilde{H}^k$ coincide.

Theorem 5.3 (Exact discrete energy balance)**.** The s-stage Gauss-Legendre methods, applied to linear PH systems with quadratic energy, are the only collocation integration schemes, which yield an exact discrete energy balance (5.43).

Proof. Equation (5.52) is, among collocation schemes, only satisfied for Gauss-Legendre collocation, see [76], Section IV.2.1, Theorem 2.2 and the paragraph below the proof[5]. If (5.52) is true, $\Delta\bar{H}^k = \Delta\tilde{H}^k$ holds, and (5.43) follows from (5.38). □

Remark 5.6. Gauss-Legendre collocation with $s = 1$ leads to the implicit midpoint rule, which is shown in [4] to satisfy an exact discrete energy balance. Theorem 5.3 shows that this is not the only choice for structure-preserving time discretization of PH systems with exact energy balance.

Remark 5.7. Beyond collocation methods, other Runge-Kutta schemes can be constructed, which satisfy the *symplecticity* condition (5.52), see [180].

5.3.2 Lobatto IIIA/IIIB Pairs

Partitioned collocation methods, such as Lobatto pairs, are used for separable Hamiltonian systems. We consider in this study the linear PH system of simple mechanical type[6] with $\mathbf{q}, \mathbf{p} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^m$,

$$\begin{aligned} \begin{bmatrix} \dot{\mathbf{q}}(t) \\ \dot{\mathbf{p}}(t) \end{bmatrix} &= \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Q}\mathbf{q}(t) \\ \mathbf{P}\mathbf{p}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{G} \end{bmatrix} \mathbf{u}(t) \\ \mathbf{y}(t) &= \begin{bmatrix} \mathbf{0} & \mathbf{G}^T \end{bmatrix} \begin{bmatrix} \mathbf{Q}\mathbf{q}(t) \\ \mathbf{P}\mathbf{p}(t) \end{bmatrix}. \end{aligned} \tag{5.53}$$

The discrete-time structure equations can be expressed as

$$\begin{bmatrix} -\mathbf{f}_q^k \\ -\mathbf{f}_p^k \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{sn} \\ -\mathbf{I}_{sn} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e}_q^k \\ \mathbf{e}_p^k \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{G}^k \end{bmatrix} \mathbf{u}^k. \tag{5.54}$$

[5] The criterion (5.52) characterizes numerical integration methods that conserve quadratic invariants. *"[A]mong all collocation and discontinuous collocation methods [...] only the Gauss methods satisfy this criterion [...]."*

[6] $\mathbf{Q} = \mathbf{K}$ denotes the stiffness matrix, $\mathbf{P} = \mathbf{M}^{-1}$ the inverse mass matrix in the common notation for linear mechanical systems.

The elements of the effort and input vectors $\mathbf{e}_q^k, \mathbf{e}_p^k \in \mathbb{R}^{sn}$, $\mathbf{u}^k \in \mathbb{R}^{sm}$ are

$$\mathbf{e}_{q,i}^k = \mathbf{Q}\mathbf{q}_i^k, \qquad \mathbf{e}_{p,i}^k = \mathbf{P}\mathbf{p}_i^k, \qquad \mathbf{u}_i^k = \mathbf{u}(t_0^k + c_i h). \tag{5.55}$$

The partitioned integration scheme, which consists of two RK methods (coefficients a_{ij}, $\hat{a}_{ij}$, $b_j = \hat{b}_j$, $c_i = \hat{c}_i$, $i, j = 1, \ldots, s$), each applied to one set of differential equations, can be written ($i = 1, \ldots, s$)

$$\mathbf{q}_0^k = \mathbf{q}_{s+1}^{k-1}, \qquad \mathbf{p}_0^k = \mathbf{p}_{s+1}^{k-1}, \tag{5.56a}$$

$$\mathbf{q}_i^k = \mathbf{q}_0^k - h\sum_{j=1}^{s} a_{ij}\mathbf{f}_{q,j}^k, \qquad \mathbf{p}_i^k = \mathbf{p}_0^k - h\sum_{j=1}^{s} \hat{a}_{ij}\mathbf{f}_{p,j}^k, \tag{5.56b}$$

$$\mathbf{q}_{s+1}^k = \mathbf{q}_0^k - h\sum_{j=1}^{s} b_j\mathbf{f}_{q,j}^k, \qquad \mathbf{p}_{s+1}^k = \mathbf{p}_0^k - h\sum_{j=1}^{s} \hat{b}_j\mathbf{f}_{p,j}^k, \tag{5.56c}$$

with $\mathbf{f}_{q,j}^k = -\mathbf{e}_{p,j}^k$, $\mathbf{f}_{p,j}^k = \mathbf{e}_{q,j}^k - \mathbf{G}^k\mathbf{u}_j^k$, $j = 1, \ldots, s$.

In contrast to Gauss-Legendre collocation, the expression for the increment of a quadratic Hamiltonian per sampling interval

$$\Delta\bar{H}^k = \frac{1}{2}\left((\mathbf{q}_{s+1}^k)^T\mathbf{Q}\mathbf{q}_{s+1}^k - (\mathbf{q}_0^k)^T\mathbf{Q}\mathbf{q}_0^k + (\mathbf{p}_{s+1}^k)^T\mathbf{P}\mathbf{p}_{s+1}^k - (\mathbf{p}_0^k)^T\mathbf{P}\mathbf{p}_0^k\right) \tag{5.57}$$

does *not* coincide with the approximate supplied energy

$$\begin{aligned}\Delta\tilde{H}^k &= -\int_{t_0^k}^{t_0^k+h} \tilde{\mathbf{e}}_q^T(s)\tilde{\mathbf{f}}_q(s) + \tilde{\mathbf{e}}_p^T(s)\tilde{\mathbf{f}}_p(s)\,ds \\ &= -h(\mathbf{e}_q^k)^T\mathbf{M}\mathbf{f}_q^k - h(\mathbf{e}_p^k)^T\mathbf{M}\mathbf{f}_p^k,\end{aligned} \tag{5.58}$$

where $\mathbf{M}$ is given by (5.26).

We restrict our attention to the 3-stage Lobatto pair[7] with collocation points $c_1 = 0$, $c_2 = \frac{1}{2}$, $c_3 = 1$, in which case the matrix $\mathbf{M}$ is

$$\mathbf{M} = \begin{bmatrix} \frac{2}{15} & \frac{1}{15} & -\frac{1}{30} \\ \frac{1}{15} & \frac{8}{15} & \frac{1}{15} \\ -\frac{1}{30} & \frac{1}{15} & \frac{2}{15} \end{bmatrix} \otimes \mathbf{I}_n. \tag{5.59}$$

Theorem 5.4. For the 3-stage Lobatto pair, applied to the partitioned linear PH system (5.53), the error between $\Delta\tilde{H}^k$ and $\Delta\bar{H}^k$, and consequently the error between $\Delta\tilde{H}^k$ and ΔH^k has order $o(h^{2s-2}) = o(h^4)$.

Proof. First notice that the local energy error $\Delta\bar{H}^k - \Delta H^k$ is of order $o(h^p)$ with $p = 2s - 2$ the order of the Lobatto pair, see Theorem 5.2. To prove that the error $\Delta\tilde{H}^k - \Delta\bar{H}^k$ has the same order, the expressions in (5.57) and (5.58)

[7] See the Butcher tables in Appendix A.2.

are subtracted, under substitution of the efforts and states in the i-th stage according to (5.55) and (5.56b). We replace the terms $\mathbf{f}_{q,1}^k$, $\mathbf{f}_{q,3}^k$ and $\mathbf{f}_{p,1}^k$, $\mathbf{f}_{p,3}^k$ by their Taylor expansions

$$\begin{aligned} \mathbf{f}_{q,1}^k &= \mathbf{f}_{q,2}^k - \dot{\mathbf{f}}_{q,2}^k \frac{h}{2} + \mathbf{r}_1 h^2, \qquad & \mathbf{f}_{q,3}^k &= \mathbf{f}_{q,2}^k + \dot{\mathbf{f}}_{q,2}^k \frac{h}{2} + \mathbf{r}_2 h^2, \\ \mathbf{f}_{p,1}^k &= \mathbf{f}_{p,2}^k - \dot{\mathbf{f}}_{p,2}^k \frac{h}{2} + \mathbf{r}_3 h^2, \qquad & \mathbf{f}_{p,3}^k &= \mathbf{f}_{p,2}^k + \dot{\mathbf{f}}_{p,2}^k \frac{h}{2} + \mathbf{r}_4 h^2, \end{aligned} \tag{5.60}$$

where $\dot{\mathbf{f}}_{q,2}^k = \frac{d}{dt}\tilde{\mathbf{f}}_q(t_2^k)$ and $\dot{\mathbf{f}}_{p,2}^k = \frac{d}{dt}\tilde{\mathbf{f}}_p(t_2^k)$ are the time derivatives of the polynomial flow approximations $\tilde{\mathbf{f}}_q$ and $\tilde{\mathbf{f}}_p$ in $t_2^k = t_0^k + \frac{h}{2}$. $\mathbf{r}_1, \ldots, \mathbf{r}_4$ are residual terms. The result is

$$\Delta\bar{H}^k - \Delta\tilde{H}^k = F_1(\mathbf{r}_1, \ldots, \mathbf{r}_4, \dot{\mathbf{f}}_{q,2}^k, \dot{\mathbf{f}}_{p,2}^k)h^5 + F_2(\mathbf{r}_1, \ldots, \mathbf{r}_4)h^6, \tag{5.61}$$

where F_1 and F_2 are functions in the given arguments. This, together with the order of the error $\Delta\bar{H}^k - \Delta H^k$, proves the claim. □

As a consequence of Theorem 5.4, the application of the 3-stage Lobatto pair to the partitioned PH system (5.53) defines a discrete-time PH system, whose discrete energy balance is consistent with the order $p = 2s - 2 = 4$ of the numerical scheme.

Theorem 5.4 shows exemplarily at the case $s = 3$ how to prove the identical consistency order of both local energy approximation errors. The numerical experiments in the following section give evidence that the corresponding order statement also holds for the 4-stage Lobatto pair.

5.4 Numerical Experiments

We illustrate the quantitative statements concerning the accuracy of the energy approximations by the numerical simulation of a linear oscillator, whose solutions can be computed, and which therefore serves as a benchmark example. First, the conservative case, which has been considered throughout the chapter, is studied. The accumulated errors of energy supplied through the port $(u(t), y(t))$ and stored energy are determined and illustrated for both considered families of integration schemes. In a second part, the control port is closed by constant feedback, which injects damping to the system. We show that the accuracy order of the energy approximation is maintained in the lossy case.

The considered state space PH model of the lossless oscillator is given by the explicit representation of the underlying Dirac structure

$$-\mathbf{f}(t) = \mathbf{J}\mathbf{e}(t) + \mathbf{g}u(t) \tag{5.62a}$$

$$y(t) = \mathbf{g}^T\mathbf{e}(t) \tag{5.62b}$$

with flow and effort vectors $\mathbf{f}, \mathbf{e} \in \mathbb{R}^2$, in- and output $u, y \in \mathbb{R}$.

$$\mathbf{J} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{g} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{5.63}$$

denote the interconnection matrix and the input vector. The dynamics equation is $\dot{\mathbf{x}}(t) = -\mathbf{f}(t)$ with $\mathbf{x} \in \mathbb{R}^2$ the state vector. The linear constitutive equations $\mathbf{e}(t) = \mathbf{Q}\mathbf{x}(t)$ are derived from the quadratic Hamiltonian $H(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$ with $\mathbf{Q} = \mathbf{I}$. For the lossy case, the extended output feedback

$$u(t) = -ry(t) + v(t), \qquad r > 0, \tag{5.64}$$

with new input $v(t)$, generates the damped system's state differential equation

$$\dot{\mathbf{x}}(t) = (\mathbf{J} - \mathbf{R})\mathbf{Q}\mathbf{x}(t) + \mathbf{g}v(t), \qquad \mathbf{R} = \begin{bmatrix} 0 & 0 \\ 0 & r \end{bmatrix}. \tag{5.65}$$

The structure equations (5.62) are discretized using collocation as described in the previous sections. First, Gauss-Legendre collocation with $s = 1, 2, 3$ stages is used. Then, the partitioned representation of (5.62) is considered for the discretization of the Dirac structure with 3- and 4-stage Lobatto pairs. Discrete-time dynamics and constitutive equations are discretized according to Definition 5.1, again considering the partitioned version of the state space model for the Lobatto pairs.

5.4.1 Energy Supply and Storage in the Lossless Case

Starting from an initial state $\mathbf{x}(0) = \begin{bmatrix} q(0) & p(0) \end{bmatrix}^T = \begin{bmatrix} 0 & -1 \end{bmatrix}^T$, the undamped system is excited by a pulse-shaped input

$$u(t) = \begin{cases} 0, & t < 8 \\ \sin^2(\frac{t-8}{10-8}\pi), & 8 \le t \le 10 \\ 0, & t > 10. \end{cases} \tag{5.66}$$

Figure 5.2 shows the exact evolutions of states, input and the quadratic energy for this test case on the interval $[0, T_{\text{end}}] = [0, 18]$. Figure 5.3 depicts the magnitudes of relative errors

$$\tilde{\epsilon}_H := \frac{\Delta\tilde{H}_{\text{tot}} - \Delta H_{\text{tot}}}{\Delta H_{\text{tot}}}, \qquad \bar{\epsilon}_H := \frac{\Delta\bar{H}_{\text{tot}} - \Delta H_{\text{tot}}}{\Delta H_{\text{tot}}} \tag{5.67}$$

of *total* supplied and stored energy over the range of step sizes $h \in [0.005, 0.5]$. $\Delta H_{\text{tot}} = \sum_{k=1}^{N} \Delta H^k$, $\Delta\tilde{H}_{\text{tot}} = \sum_{k=1}^{N} \Delta\tilde{H}^k$ and $\Delta\bar{H}_{\text{tot}} = \sum_{k=1}^{N} \Delta\bar{H}^k$ denote the total increment of energy and its approximations on $[0, T_{\text{end}}]$ with $N = T_{\text{end}}/h$ the number of sampling intervals. With $\Delta\bar{H}^k = \Delta H^k + c^k h^{p+1}$, where p is the order of the integration scheme, the absolute value of $\bar{\epsilon}_H$ can be bounded as follows:

$$|\bar{\epsilon}_H| = \frac{|\sum_{k=1}^{N} c^k h^{p+1}|}{|\sum_{k=1}^{N} \Delta H^k|} \le \frac{\max_k |c^k|}{|P_{\text{av}}|} h^p. \tag{5.68}$$

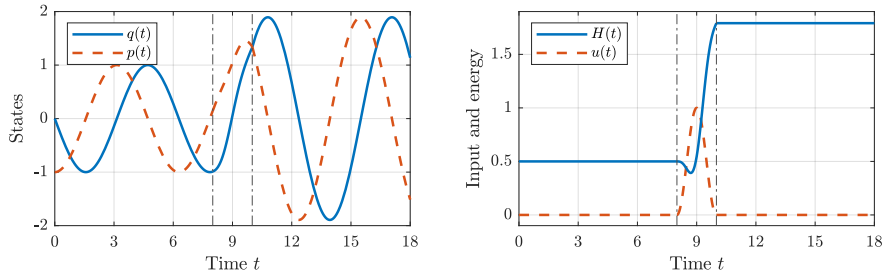

Figure 5.2: Evolution of states (left), input and energy (right) for the forced undamped oscillator.

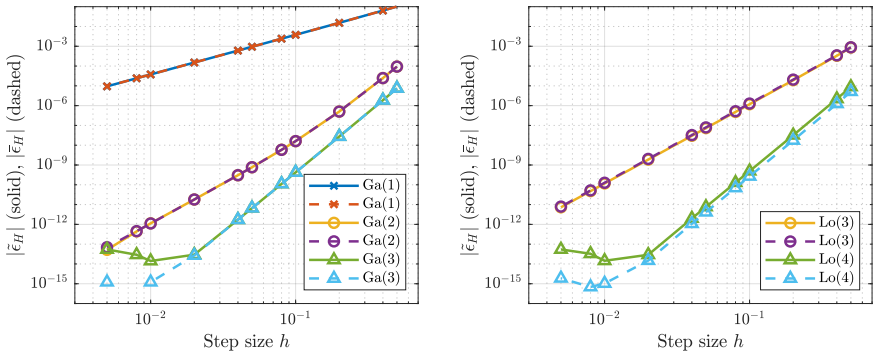

Figure 5.3: Errors $\tilde{\epsilon}_H$ and $\bar{\epsilon}_H$ of supplied and stored energy for Gauss-Legendre methods, $s = 1, 2, 3$ (left) and Lobatto pairs, $s = 3, 4$ (right).

$P_{\text{av}} = \frac{\Delta H_{\text{tot}}}{Nh}$ denotes the average transferred power, and the same estimation of order can be performed for $\tilde{\epsilon}_H$.

The first diagram in Figure 5.3 nicely shows the orders 2, 4 and 6 of the Gauss-Legendre methods as well as the fact that both approximations $\Delta\tilde{H}^k$ and $\Delta\bar{H}^k$ of supplied and stored energy coincide. (The effect of rounding errors becomes visible at low step sizes in the curve for $s = 3$.) The slopes in the second diagram confirm the orders 4 and 6 of the 3- and 4-stage Lobatto pairs. Although hardly recognizable, the two curves for $\tilde{\epsilon}_H$ and $\bar{\epsilon}_H$ do not match, which is accordance with the computations for $s = 3$, see Eq. (5.61).

5.4.2 Approximation of Dissipated Energy

The damped oscillator represents the most basic power-conserving interconnection of a PH system (the undamped oscillator) with another system (a purely resistive element). This is nicely seen if (5.62) is combined with the damping injection feedback (5.64). The differential energy balance in this damped case

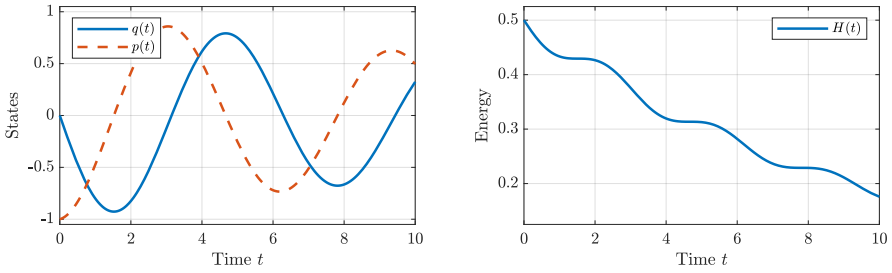

Figure 5.4: Evolution of the states (left) and the energy (right) for the damped oscillator.

reads

$$\dot{H}(t) = -ry^2(t) + y(t)v(t) \leq y(t)v(t), \tag{5.69}$$

which is the balance of power to the energy storage elements, supplied power and dissipated power. In the sequel, we set $v(t) \equiv 0$. Figure 5.4 shows the solution of (5.65) with damping parameter $r = 0.1$ and an initial value $\mathbf{x}(0) = \begin{bmatrix} p(0) & q(0) \end{bmatrix}^T = \begin{bmatrix} 0 & -1 \end{bmatrix}^T$ on the time interval $[0, 10]$, as well as the monotonous decrease in energy.

Discretization of the damped PH model, evaluation of the total dissipated energy ΔH_{tot} and comparison with the numerical values $\Delta \bar{H}_{\text{tot}}$ and $\Delta \tilde{H}_{\text{tot}}$ results in the error plots in Fig. 5.5. As in the lossless case, the error plots confirm that the dissipated power is discretized consistently with the order of the underlying geometric numerical integration scheme. This time, the discrepancy between the numerical energy increments $\Delta \tilde{H}^k$ and $\Delta \bar{H}^k$ for the Lobatto pairs, which is of order $\mathcal{O}(h^p)$, is clearly visible in the right diagram.

Remark 5.8. The discrete-time correspondence of the damping injection feedback

$$u(t) = -ry(t) = -rx_2(t), \tag{5.70}$$

which produces the discrete-time PH system with dissipation is

$$u_i^k = -rx_{2,i}^k, \qquad i = 1, \ldots, s, \tag{5.71}$$

and *not* $\mathbf{u}^k = -r\mathbf{y}^k$ with the output $\mathbf{y}^k$ as defined in (5.31). This point is very clear, when we consider the different meanings of the ports $(u(t), y(t))$ and $(\mathbf{u}^k, \mathbf{y}^k)$: The pairing of the former port variables gives the *instantaneous* power, while the latter approximates the *average* transmitted power over the sampling interval I^k.

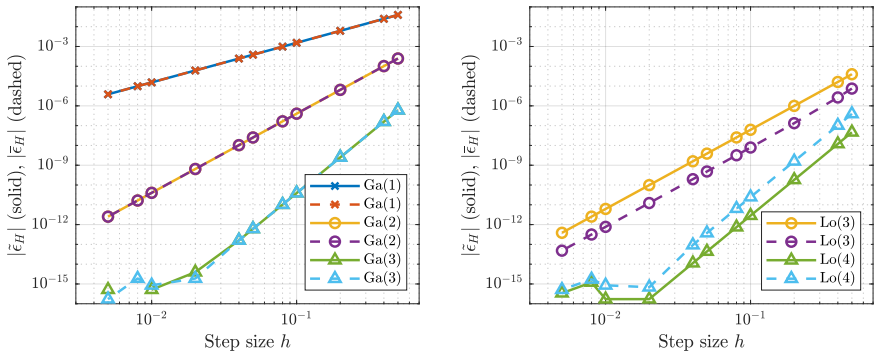

Figure 5.5: Errors $\tilde{\epsilon}_H$ and $\bar{\epsilon}_H$ of energy flow over the dissipative port and energy loss in the storage elements for Gauss-Legendre methods, $s = 1, 2, 3$ (left) and the Lobatto pair, $s = 3$ (right).

5.5 Conclusions

We presented a new definition of discrete-time PH systems, which is based on the approximation of the structure equations and the energy balance of explicit PH systems using the collocation method. By defining a discrete-time Dirac structure, discrete-time constitutive equations and using appropriate geometric numerical integration schemes, the separation between structure, constitutive laws and dynamics, which is a central feature of PH systems, is maintained. The presented work extends in a very natural way the notion of geometric/symplectic integration of autonomous Hamiltonian systems to the *open* case (i. e. with power flow over the system boundary) of PH systems.

A focus has been set on proving consistency of the two different numerical energy increments that appear in the context of this definition. The family of implicit Gauss-Legendre schemes – applied to linear PH systems – is the only one among collocation schemes for which the approximations of supplied and stored energy match, which leads to an *exact* discrete energy balance of the discrete-time PH approximation. For Lobatto IIIA/IIIB pairs, applied to a linear PH system of partitioned mechanical structure, the discrete energy balance is not exact, but the energy error is consistent with the numerical integration scheme. The theoretical findings have been illustrated by numerical experiments with the simplest test case of a linear oscillator: The evolution of total energy, which is (i) supplied by an external input or (ii) dissipated based on output damping injection, is approximated up to the order of the underlying integration scheme.

The presented definition of discrete-time PH systems can be exploited in the simulation and numerical analysis of large scale networks. The consistent approximation of energy flows between subsystems and the quantification of their errors give important insight that helps to keep track of the quality of

simulation results. In the context of network simulation, the numerical approximation of PH DAE systems [190], [13] is of particular interest. Combined with structure-preserving spatial discretization, see e. g. [104], the presented approach contributes to the full discretization of distributed parameter PH systems. Moreover, the presented work gives rise to reconsider the *Control by Interconnection* approach, see e. g. [149], for the stabilization of PH systems in discrete time. Finally, more general choices for the time discretization of effort and flow variables are conceivable, which would lead to interesting implicit representations of Dirac structures with implicit discrete dynamics.

Chapter 6

Preservation of Flatness and Feedforward Control

Differential flatness, see e. g. the article [60] and the books [176], [116], is a system property, which is extremely useful for feedforward and feedback control. In continuous-time finite-dimensional *flat* systems, state and input trajectories can be computed *algebraically* from a *flat output* and a finite number of its time derivatives. This means that the system dynamics can be inverted without integration, or in other words, the occurrence of internal dynamics. In the linear case, controllability guarantees the existence of a flat output, which has full relative degree. For discrete-time finite-dimensional systems, flatness can be defined in an analogous way, see [176], Chapter 5. The *differential parametrization* of states and inputs is replaced by a *difference parametrization*, i. e. states and inputs can be expressed in terms of the output sequence and a finite number of forward/backward shifts in discrete time. As for the continuous-time case [117], the characterization of differentially flat nonlinear systems in discrete time [91] is a non-trivial task.

To extend the flatness approach to infinite-dimensional systems, there exists a series of approaches with different properties depending on the class of infinite-dimensional systems. For a tubular reactor [63] or the one-dimensional heat equation [109], a flat parametrization can be obtained by the *ansatz* of an infinite power series for the state. Convergence of the state and the input power series (to be truncated for implementation) depends on the smoothness of the desired output trajectory, which is given for Gevrey class [68] functions. Alternatively, the *Riesz spectral property* of the linear system operator can be exploited, also on higher-dimensional spatial domains, which also leads to power series representations of the system variables [14], [133], [134]. Another approach is to use *Mikusiński's operational calculus*, see e. g. [62], [121] and the books [166], [167] with references and examples from different physical domains. The linear PDE is transformed to an ordinary differential equation in space and solved under the boundary condition of the desired flat output. The back-transformation to time domain results in infinite power series and/or finite

distributed delays and predictions, depending on the parabolic/hyperbolic nature of the problem. Flatness and flatness-based control of hyperbolic systems [61], [93], [210], [211], where initial and boundary conditions are transported along the characteristic curves, is closely related to flatness of delay systems [139].

In order to avoid operational calculus and to enable the flatness-based feedforward control of *nonlinear* infinite-dimensional systems, the spatial discretization of the underlying PDEs is considered in [143]. Based on finite difference approximations of the heat equation and a nonlinear flexible beam, the convergence of the computed solutions under grid refinement is shown. The article [19] shows the application of [133] to temperature control of a deep-drawing process on a complex-shaped geometry using finite element models and model order reduction.

In this chapter[1], which starts with a brief summary of flatness definitions in finite dimension, we derive conditions on the parameters of the structure-preserving discretization scheme presented in Chapter 4, under which flatness of a given output is conserved. We restrict ourselves to the 1D case and perform numerical experiments to assess the quality of the resulting feedforward controllers. In continuous time, increasing the order of the numerical approximation model by grid refinement increases also the required smoothness degree of the desired output trajectory. For the *parabolic* heat equation, which is treated in Section 6.2, this is no restriction, as flatness-based feedforward control on the PDE model also requires infinitely smooth desired output functions. For *hyperbolic* systems whose solutions transport initial and boundary conditions with finite speed, smoothness up to a certain degree represents an unphysical constraint on the output trajectory. In Section 6.3, we show for the 1D wave equation how flatness is preserved, if structure-preserving spatial discretization is combined with symplectic time integration as discussed in Chapter 5. The result is flatness of the output in discrete time, which allows for a discrete-time flat parametrization of the states and the input. The computed control depends on the output and its forward/backward shifts in discrete time. The results from the linear case are extended to a class of nonlinear conservation laws in Section 6.4. The consistent approximation of constitutive equations, together with geometric time integration, allows again for flatness-based discrete-time controller design.

In Sections 6.2 and 6.3, we compare the numerical flatness-based boundary controls with the solutions that are obtained from the infinite-dimensional models, i. e. based on *late lumping* or the exact solution. Our algorithm for nonlinear conservation laws is applied to the flow routing problem of the 1D shallow water equations in Section 6.4. We validate our results against those obtained in [93], where the feedforward control is computed using the method of characteristics.

[1] Section 6.2 is based on the results of the corresponding parts of [96] and [97]. The results of Section 6.4 are published in [98].

Remark 6.1. The finite-dimensional models, for which conservation of flatness is examined in this chapter, result from a discretization scheme that preserves the port-Hamiltonian structure, in particular *passivity* of the system representation. Passivity and flatness are *a priori* two very different system properties. In the finite-dimensional case, the different relative degrees of a passive and a flat output give evidence of this. However, it is also clear that the system differential equation of a PH system can be complemented by an output which has the flatness property and whose conservation under numerical approximation is of interest for numerical control design.

6.1 Definitions

We recall the definition of flatness for finite-dimensional continuous-time and discrete-time systems as a prerequisite for the discussion in the following sections. We restrict ourselves to the SISO case of a single in- and output. For flatness of different classes of infinite-dimensional systems, we refer to the references listed in the previous paragraph.

6.1.1 Continuous-Time Systems

As introduced in [60], a nonlinear system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, u), \qquad \mathbf{x} \in \mathbb{R}^n, \quad u \in \mathbb{R}, \tag{6.1}$$

which can be feedback linearized by an *endogeneous* dynamic feedback[2] is called *(differentially) flat*. In particular, there exists a so-called *flat output*, which depends on only the system states, the input and the time derivatives of the input, such that the states and the input can be parametrized by the flat output and its time derivatives. The following definition summarizes the characterization of flatness in the introduction of [60], see also [176], Definition 7.2.1, and [116], Corollary 6.2.

Definition 6.1. The system (6.1) is called *(differentially) flat*, if there exists a function – a *flat output* –

$$y = \phi(\mathbf{x}, u, \dot{u}, \ddot{u}, \ldots, \overset{(\alpha)}{u}), \tag{6.2}$$

which allows for a *differential parametrization*

$$\mathbf{x} = \boldsymbol{\psi}_x(y, \dot{y}, \ddot{y}, \ldots, \overset{(\beta)}{y}), \qquad u = \psi_u(y, \dot{y}, \ddot{y}, \ldots, \overset{(\beta+1)}{y}) \tag{6.3}$$

of the state and the input in terms of the flat output and its time derivatives.

[2] *Endogeneous* means that *"the dynamic extension does not contain* exogeneous *variables, which are independent of the original system variables and their derivatives"*, see [60], Section 1 or [116], Subsection 5.3.6.

The implication for feedforward control design is evident: Given a desired output trajectory $y_d(t)$, which is at least $\beta+1$ times continuously differentiable, the corresponding state and input trajectories $\mathbf{x}_d(t)$ and $u_d(t)$ follow directly from substitution in (6.3).

For linear SISO systems in state space form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \qquad \mathbf{x} \in \mathbb{R}^n, \quad u \in \mathbb{R}, \tag{6.4}$$

the existence of a flat output is equivalent to controllability. The following theorem rephrases Proposition 2.4.1 in [176].

Theorem 6.1. The flat output of a SISO system (6.4) is given by

$$y = c \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix} \mathbf{Q}_c^{-1} \mathbf{x}, \qquad c \in \mathbb{R}\backslash\{0\}, \tag{6.5}$$

where

$$\mathbf{Q}_c = \begin{bmatrix} \mathbf{b} & \mathbf{A}\mathbf{b} & \dots & \mathbf{A}^{n-1}\mathbf{b} \end{bmatrix} \tag{6.6}$$

denotes Kalman's controllability matrix.

As a consequence of this theorem, a given scalar output

$$y = \mathbf{c}^T \mathbf{x} \tag{6.7}$$

of the LTI system (6.4) is flat if and only if the output vector $\mathbf{c}^T \in \mathbb{R}^{1\times n}$, multiplied with the controllability matrix $\mathbf{Q}_c$, gives the n-th unit vector modulo a constant factor:

$$\mathbf{c}^T \mathbf{Q}_c = \frac{1}{c} \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix}. \tag{6.8}$$

This requirement can be recast as

$$\mathbf{c}^T \mathbf{A}^k \mathbf{b} = 0, \qquad k = 0, \dots, n-2, \tag{6.9a}$$

$$\mathbf{c}^T \mathbf{A}^{n-1} \mathbf{b} \neq 0. \tag{6.9b}$$

Moreover, to express $\mathbf{x}$ only in terms of $y, \dot{y}, \dots, \overset{(n-1)}{y}$ (see [176], Section 2.4), the output (6.7) must be observable.

Corollary 6.1. A flat output of a controllable LTI system (6.4) is observable and has full relative degree n.

6.1.2 Discrete-Time Systems

Flatness of controllable discrete-time systems

$$\mathbf{x}^{k+1} = \mathbf{A}_d \mathbf{x}^k + \mathbf{b}_d u^k \tag{6.10}$$

is discussed in Chapter 5 of [176] along the same lines as for the continuous-time case. Controllability of the discrete-time state space model guarantees the transformation $\mathbf{z} = \mathbf{Q}_{c,d}^{-1}\mathbf{x}$ to *controllability* normal form[3], where $\mathbf{Q}_{c,d}$ is the discrete-time controllability matrix. It can be easily verified that the last transformed state z_n represents a flat output for (6.10), see Proposition 5.4.1 in [176].

Theorem 6.2. The flat output of a discrete-time SISO system (6.10) is given by

$$y = c\begin{bmatrix}0 & \dots & 0 & 1\end{bmatrix}\mathbf{Q}_{c,d}^{-1}\mathbf{x}, \qquad c \in \mathbb{R}\backslash\{0\}, \tag{6.11}$$

where

$$\mathbf{Q}_{c,d} = \begin{bmatrix}\mathbf{b}_d & \mathbf{A}_d\mathbf{b}_d & \dots & \mathbf{A}_d^{n-1}\mathbf{b}_d\end{bmatrix} \tag{6.12}$$

denotes Kalman's discrete-time controllability matrix.

The state vector $\mathbf{x}^k$ and the input u^k can be parametrized by the flat output y^k and a finite number of its predictions y^{k+l}, $l = 1, \dots, n$.

In Sections 6.3 and 6.4, we will use a *symplectic* integration scheme in order to obtain fully discretized (i. e. in space and time) state space models of hyperbolic systems. As discussed in Chapter 5, symplectic integration allows to conserve structural properties of a finite-dimensional port-Hamiltonian system (a structural energy balance, energy conservation in the zero input case) in the discrete-time model. In contrast to (6.10), the discrete-time state space models will be *implicit*. Accordingly, the flat parametrization of states and outputs will contain *predictions and delays* of the flat output.

6.2 One-Dimensional Heat Equation

Recall the structured representation of the heat equation from Subsection 2.3.3. We consider, as in Subsection 4.6.6, the case of a one-dimensional spatial domain $\Omega = (0, 1) \subset \mathbb{R}$ and constant material parameters $\lambda = c_v = 1$:

$$\begin{bmatrix}f^p\\f^q\end{bmatrix} = \begin{bmatrix}0 & \mathrm{d}\\ \mathrm{d} & 0\end{bmatrix}\begin{bmatrix}e^p\\e^q\end{bmatrix} \qquad \text{(Structure)} \tag{6.13a}$$

$$\dot{p} = -f^p \qquad \text{(Dynamics)} \tag{6.13b}$$

$$\begin{bmatrix}e^p\\e^q\end{bmatrix} = \begin{bmatrix}*p\\-*f^q\end{bmatrix} \qquad \text{(Constit. Eq.)} \tag{6.13c}$$

$$\begin{bmatrix}0\\u\end{bmatrix} = \begin{bmatrix}e^q(0)\\e^p(1)\end{bmatrix} \qquad \text{(Boundary cond.)} \tag{6.13d}$$

$$y = e^p(0) \qquad \text{(Flat output)} \tag{6.13e}$$

[3] Not to be confused with the *controller* canonical form.

In the structure equation, $f^p, f^q \in L^2\Lambda^1(\Omega)$ represent the negative time derivative of internal energy density and the thermodynamic driving force, while $e^p, e^q \in H^1\Lambda^0(\Omega)$ denote temperature and heat flux[4]. The constitutive equations represent the calorimetric equation and Fourier's law. For $z = 0$, an insulating boundary condition (zero heat flux) is imposed, while at $z = 1$, the temperature plays the role of an input. The temperature at $z = 1$ is a flat output.

6.2.1 Feedforward Control Based on the PDE Model

We summarize the flatness-based feedforward control design for the heat equation according to [109]. As an alternative to the formal power series *ansatz* of the solution, we can start from the irrational transfer function for the system (6.13), see Appendix B.2. Having no zero, we can invert the transfer function,

$$\hat{u}(s) = \cosh(\sqrt{s})\hat{y}(s), \tag{6.14}$$

which allows to compute the input signal without integration in time. The hyperbolic cosine is expressed as an infinite series

$$\cosh(\sqrt{s}) = \sum_{j=0}^{\infty} \frac{s^j}{(2j)!}. \tag{6.15}$$

Substitution in (6.14) and backtransformation to time domain (the star indicates the *desired* output function and the corresponding input),

$$u^*(t) = \sum_{j=0}^{\infty} \frac{1}{(2j)!} \frac{d^j}{dt^j} y^*(t), \tag{6.16}$$

yield the flat parametrization of the temperature at the left boundary (input) by the temperature at the right boundary (flat output).

To realize a transient between two stationary outputs $y^*(0) = 0$ and $y^*(T) = 1$, the desired trajectory must be infinitely often differentiable on $[0, T]$, while all time derivatives must be zero in $t = 0$ and $t = T$. This means that at these two points, $y^*(t)$ must be non-analytic [135]. The smoothened step function $y^*(\tau) = \Theta_\omega(\tau)$, $\tau = \frac{t}{T}$,

$$\Theta_\omega(\tau) = \begin{cases} 0 & \tau \leq 0, \\ \frac{\int_0^\tau \theta_\omega(s)ds}{\int_0^1 \theta_\omega(s)ds} & \tau \in (0,1), \\ 1 & \tau \geq 1, \end{cases} \tag{6.17}$$

[4]Note that the functional spaces are defined based on only the structure equation, which is discretized in a first step.

satisfies the above-mentioned conditions. $\theta_\omega(\tau)$ represents the "bump" function

$$\theta_\omega(\tau) = \begin{cases} 0 & \tau \notin (0,1) \\ \exp\left(-[(1-\tau)\tau]^{-\omega}\right) & \tau \in (0,1). \end{cases} \tag{6.18}$$

For a parameter choice $\omega > 1$, the desired trajectory $y^*(\tau)$ is a *Gevrey class* function of order[5] $1 < \gamma < 2$ with $\gamma = 1 + \frac{1}{\omega}$. A value of $\gamma < 2$ guarantees convergence of the power series (6.16) with infinite convergence radius [120]. To compute the input trajectory $u^*(t)$, the infinite series (6.16) is truncated after a finite number of elements.

6.2.2 Feedforward Control Based on the Discretization

The structure-preserving discretization of the PDE model (6.13) according to Subsection 4.6.6 yields the SISO linear state space model (4.119), (4.121) of order N. We consider the two cases $\alpha = 0$ and $\alpha = \frac{1}{2}$, which correspond to a one-sided and a centered approximation of the constitutive equations, respectively.

Case 1, $\alpha = 0$

The state space model $(\mathbf{A}, \mathbf{b}, \mathbf{c}^T)$ has no invariant zeros, i. e. the numerator polynomial of the transfer function

$$\frac{Y(s)}{U(s)} = \mathbf{c}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \frac{b_0}{s^N + a_{N-1}s^{N-1} + \ldots + a_1 s + a_0} \tag{6.19}$$

is a constant. Consequently, the input $U^*(s)$ is obtained by multiplication of the desired output $Y^*(s)$ with the N-th order characteristic polynomial:

$$U^*(s) = \frac{s^N + a_{N-1}s^{N-1} + \ldots + a_1 s + a_0}{b_0} Y^*(s). \tag{6.20}$$

In the time domain, this corresponds to a weighted sum of $y^*(t)$ and its time derivatives up to order N:

$$u^*(t) = \frac{1}{b_0} \sum_{j=0}^{N} a_j \frac{d^j}{dt^j} y^*(t), \qquad a_N = 1. \tag{6.21}$$

Case 2, $\alpha = \frac{1}{2}$

Assuming an even number of discretization intervals N, the discretized state space model $(\mathbf{A}, \mathbf{b}, \mathbf{c}^T)$ can be split into two subsystems, see (4.139). The second subsystem is unobservable due to the compensation of exactly one half of

[5] We use the symbol γ for the order of the Gevrey class function (instead of α in [68]) to avoid confusion with our mapping parameter α.

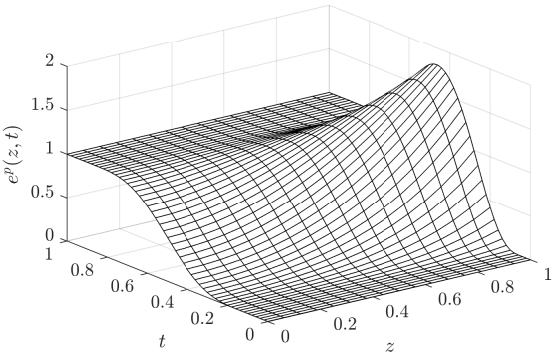

Figure 6.1: Simulation result with the feedforward controller for an s-shaped transient of $y = T(0) = e^p(0)$ with $\omega = 1.1$. Controller design: $N = 40$, $\alpha = \frac{1}{2}$, using y and its time derivatives up to order 10. Simulation: $N_{sim} = 160$, $\alpha = \frac{1}{2}$.

the eigenvalues (of multiplicity 2) with the invariant zeros. The first subsystem is controllable and observable and has a transfer function of the form

$$\begin{aligned} \frac{Y(s)}{U(s)} &= \mathbf{c}_1^T (s\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{b}_1 \\ &= \frac{b'_0}{s^{\frac{N}{2}} + a'_{\frac{N}{2}-1} s^{\frac{N}{2}-1} + \ldots + a'_1 s + a'_0}, \end{aligned} \tag{6.22}$$

from which, in analogy to above, the flat parametrizaion of the input can be obtained:

$$u^*(t) = \frac{1}{b'_0} \sum_{j=0}^{\frac{N}{2}} a'_j \frac{d^j}{dt^j} y^*(t), \qquad a'_{\frac{N}{2}} = 1. \tag{6.23}$$

Remark 6.2. Note that for the discretization with $\alpha = \frac{1}{2}$, the given output $y(t)$ can be considered a flat output of the controllable and observable subsystem $(\mathbf{A}_1, \mathbf{b}_1, \mathbf{c}_1^T)$. The unobservable subsystem $(\mathbf{A}_2, \mathbf{b}_2, \mathbf{c}_2^T)$, which has the same negative real eigenvalues as the first one, is excited by the input $u(t)$. Its solution $\mathbf{x}_2(t)$ tends asymptotically to an equilibrium depending on the steady state value of $u(t)$ and does not affect the state $\mathbf{x}_1(t)$ of the first subsystem.

6.2.3 Numerical Experiments

We use a simulation model with $\alpha = \frac{1}{2}$, $N_{sim} = 160$, which is integrated using Matlab's lsim with a time step of 10^{-5}. We consider a transient reference output between $y(0) = 0$ and $y(1) = 1$, which is described by (6.17). With the parameter $\omega = 1.1$, the output trajectory is of Gevrey order $1 + \frac{1}{1.1} < 2$. The infinite series (6.16) to compute the flat input parametrization is truncated

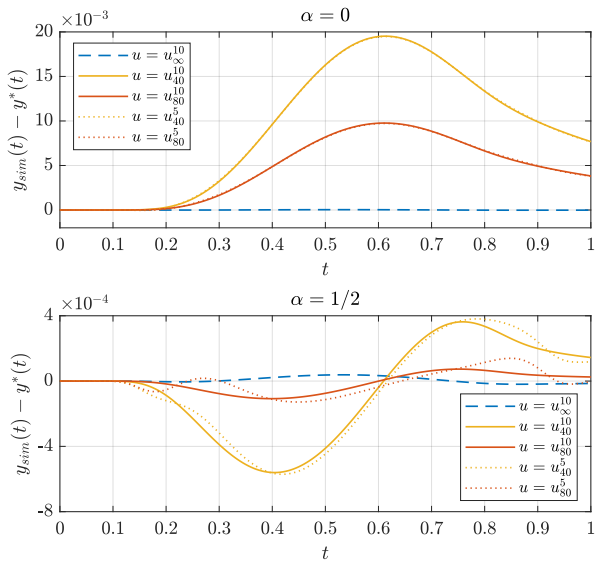

Figure 6.2: Output errors under flatness-based feedforward control. Simulation model: $N_{sim} = 160$, $\alpha = \frac{1}{2}$. u_c^{10}: Analytic computation of feedforward control using 10 time derivatives of y. $u_{40/80}^{5/10}$: Computation of feedforward control based on the discretized model with $\alpha = \frac{1}{2}$, $N \in \{40, 80\}$ using 5 or 10 time derivatives of $y^*(t)$.

after the 10th time derivative of $y^*(t)$. The corresponding controller is denoted $u_\infty^{10}(t)$.

For comparison, we compute the feedforward control based on (6.21) for $\alpha = 0$ and (6.23) for $\alpha = \frac{1}{2}$. Both series are also truncated after the first 10 (or 5, respectively) time derivatives. The corresponding controllers are denoted $u_N^{5/10}(t)$, where the subscript denotes the order N of the control design model and the superscript the number of time derivatives of the flat output that were used.

Figure 6.1 illustrates the vector of lumped co-energy variables (nodal temperatures) $\mathbf{e}^p(t) = \mathbf{Q}_p\tilde{\mathbf{p}}(t) = \mathbf{Q}_p\mathbf{x}(t)$ over space and time for the temperature transient with input $u_{40}^{10}(t)$, imposed at $z = 1$. In Fig. 6.2, we compare the simulated output with the reference trajectory for $\alpha = 0$ (top) and $\alpha = \frac{1}{2}$ (bottom), under variation of $N \in \{40, 80\}$ and the number of used output derivatives (10 or 5). The different order (1 vs. 2) of the approximation error is evident from the curves: The magnitude of the output error is reduced to a factor $\frac{1}{2}$ and $\frac{1}{4}$, respectively, when doubling the number of discretization intervals, which is consistent with the analysis of the simulation models in Subsection 4.6.6, Fig. 4.8. Comparing the solid with the dotted curves, the decreasing influence of higher order derivatives of the flat output can be observed. Despite the fact that for $\alpha = \frac{1}{2}$ (centered approximation), $y(t)$ is only the flat output of

the controllable and observable subsystem (the unobservable subsystem being asymptotically stable), the feedforward control based on this finite-dimensional model is superior to the case $\alpha = 0$ (one-sided approximation), in which $y(t)$ remains the flat output of the overall system.

6.3 One-Dimensional Wave Equation

In analogy to the previous section, we rewrite the structured representation of the 1D wave equation on $\Omega = (0, 1)$ with constant propagation speed $c = 1$, as introduced in Subsections 2.3.1 and 4.6.5:

$$\begin{bmatrix} f^p \\ f^q \end{bmatrix} = \begin{bmatrix} 0 & \mathrm{d} \\ \mathrm{d} & 0 \end{bmatrix} \begin{bmatrix} e^p \\ e^q \end{bmatrix} \qquad \text{(Structure)} \tag{6.24a}$$

$$\begin{bmatrix} \dot{p} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} -f^p \\ -f^q \end{bmatrix} \qquad \text{(Dynamics)} \tag{6.24b}$$

$$\begin{bmatrix} e^p \\ e^q \end{bmatrix} = \begin{bmatrix} *p \\ *q \end{bmatrix} \qquad \text{(Constit. Eq.)} \tag{6.24c}$$

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} e^q(0) \\ e^p(1) \end{bmatrix} \qquad \text{(Boundary cond.)} \tag{6.24d}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -e^q(1) \\ e^p(0) \end{bmatrix} \qquad \text{(Non-coll. outputs)} \tag{6.24e}$$

Flow and effort differential forms are $f^p, f^q \in L^2\Lambda^1(\Omega)$, $e^p, e^q \in H^1\Lambda^0(\Omega)$. Note that the outputs are defined in a *non-collocated* way[6], i. e. u_1 and y_1 represent the same physical quantity at opposite boundaries, accordingly for u_2 and y_2.

We address the problem of finding the *upstream* input $u_1^*(t)$ for a given *downstream* output trajectory $y_1^*(t)$ under an algebraic boundary condition

$$B(u_2, y_1) = B(e^p(1), -e^q(1)) = 0 \tag{6.25}$$

for the effort variables at $z = 1$. For the linear wave equation, we make the particular choice of

$$u_2 = -y_1, \tag{6.26}$$

which corresponds for example to the perfectly absorbing resistive termination of a transmission line, see the example[7] in [71].

[6] (u_1, y_2) and (u_2, y_1) represent the (collocated) boundary *ports*.

[7] Therein, the transmission line of length $e - 1$ has identical distributed inductance and capacitance functions that depend on z. The line is terminated with a lumped resistance of value 1. The exact solution of the underlying PDE, as in our problem with constant parameters, yields simply $y_1(t) = u_1(t - 1)$.

6.3.1 Solution of the PDE Model

In classical notation, with $p(z,t)$ and $q(z,t)$ functions in space and time instead of differential forms, the above-described (inverse) boundary control problem reads: Find

$$u_1^*(t) = q(0,t) \qquad (\text{and } y_2^*(t) = p(0,t)), \tag{6.27}$$

which satisfy the partial differential equations on $\Omega = (0,1)$

$$\begin{aligned} \tfrac{\partial}{\partial t} p(z,t) &= -\tfrac{\partial}{\partial z} q(z,t) \\ \tfrac{\partial}{\partial t} q(z,t) &= -\tfrac{\partial}{\partial z} p(z,t) \end{aligned} \tag{6.28}$$

under the boundary conditions

$$\begin{aligned} q(1,t) &= -y_1^*(t) \\ p(1,t) &= u_2^*(t) = -y_1^*(t). \end{aligned} \tag{6.29}$$

In new coordinates $\xi(z,t) = \frac{1}{2}(p(z,t)+q(z,t))$ and $\eta(z,t) = \frac{1}{2}(p(z,t)-q(z,t))$, the PDEs are decoupled transport equations

$$\begin{aligned} \tfrac{\partial}{\partial t} \xi(z,t) &= -\tfrac{\partial}{\partial z} \xi(z,t) \\ \tfrac{\partial}{\partial t} \eta(z,t) &= \tfrac{\partial}{\partial z} \eta(z,t) \end{aligned} \tag{6.30}$$

and the boundary conditions transform to

$$\begin{aligned} \xi(1,t) &= -y_1^*(t) \\ \eta(1,t) &= 0. \end{aligned} \tag{6.31}$$

Functions $\xi(z,t) = f_\xi(z-t)$, $\eta(z,t) = f_\eta(z+t)$, which are given in the *characteristic variables* $z-t$ and $z+t$, solve the PDEs (6.30), and represent a right- and a left-travelling wave, respectively. Taking into account the boundary conditions (6.31), finally yields the solution

$$\begin{aligned} \xi(z,t) &= -y_1^*(t+1-z) \\ \eta(z,t) &= 0 \end{aligned} \tag{6.32}$$

or

$$\begin{aligned} p(z,t) &= -y_1^*(t+1-z) \\ q(z,t) &= -y_1^*(t+1-z). \end{aligned} \tag{6.33}$$

The corresponding boundary input

$$u_1^*(t) = q(0,t) = -y_1^*(t+1) \tag{6.34}$$

is the desired output trajectory (up to the minus sign due to the power flow convention), advanced by the transport delay 1. This solution serves as a reference for the study of feedforward controllers based on finite-dimensional numerical approximations of the wave equation.

6.3.2 Semi-Discretization

Structure-preserving spatial discretization of the wave equation (6.24) as described in Subsection 4.6.5 yields the linear continuous-time state space model[8]

$$\begin{bmatrix}\dot{\tilde{\mathbf{p}}}\\ \dot{\tilde{\mathbf{q}}}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{J}_p\\ -\mathbf{J}_p^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_q\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{p}}\\ \tilde{\mathbf{q}}\end{bmatrix} + \begin{bmatrix}\mathbf{B}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{B}_q\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix}, \tag{6.35a}$$

$$\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{B}_q^T\\ \mathbf{B}_p^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_q\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{p}}\\ \tilde{\mathbf{q}}\end{bmatrix}. \tag{6.35b}$$

The matrices on the right hand side are given by (4.102) and (4.106). Invoking the boundary condition (6.26), and disregarding y_2, we obtain the SISO linear system

$$\begin{bmatrix}\dot{\tilde{\mathbf{p}}}\\ \dot{\tilde{\mathbf{q}}}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{J}_p\\ -\mathbf{J}_p^T & -\mathbf{B}_q\mathbf{B}_q^T\end{bmatrix}\begin{bmatrix}\mathbf{Q}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_q\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{p}}\\ \tilde{\mathbf{q}}\end{bmatrix} + \begin{bmatrix}\mathbf{B}_p\\ \mathbf{0}\end{bmatrix}u_1, \tag{6.36a}$$

$$y_1 = \begin{bmatrix}\mathbf{0} & \mathbf{B}_q^T\end{bmatrix}\begin{bmatrix}\mathbf{Q}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_q\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{p}}\\ \tilde{\mathbf{q}}\end{bmatrix}. \tag{6.36b}$$

Theorem 6.3. The output y_1 of the linear SISO system (6.36), which results from the structure-preserving spatial discretization of the wave equation (6.24), is only a flat output for the choice of the discretization parameter $\alpha = 0$.

Proof. For the considered state space model of order $n = 2N$, $\mathbf{c}^T\mathbf{b} = 0$ and the terms $\mathbf{c}^T\mathbf{A}^k\mathbf{b}$, $k = 1, \ldots, n-2$ are polynomials and rational functions in α, whose only common root is 0. Hence, condition (6.9a) is only satisfied for $\alpha = 0$. With $\mathbf{c}^T\mathbf{A}^{n-1}\mathbf{b} \neq 0$, the output y_1 has full relative degree n. The system matrices with $\alpha = 0$ are

$$\mathbf{A} = \frac{1}{h}\begin{bmatrix} & & & & -1 & & & \\ & & & & 1 & -1 & & \\ & & & & & \ddots & \ddots & \\ & & & & & & 1 & -1\\ 1 & -1 & & & & & & \\ & \ddots & \ddots & & & & & \\ & & 1 & -1 & & & & \\ & & & 1 & & & & -1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix}1\\0\\ \vdots \\0\end{bmatrix}, \tag{6.37}$$

$$\mathbf{c}^T = \frac{1}{h}\begin{bmatrix}0 & \ldots & 0 & -1\end{bmatrix}.$$

Controllability can be verified based on the structure of Kalman's controllability matrix. Collecting first the odd columns, then the even ones, the controllability

[8] Note the permutation of the outputs compared to (4.108), which are now non-collocated with respect to the inputs.

matrix with permuted columns has the structure

$$\mathbf{Q}_{c,perm} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}, \tag{6.38}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are upper triangular matrices with non-zero diagonal elements. Observability of y_1 can be shown in a similar way, which completes the proof. □

Remark 6.3. The LTI system $(\mathbf{A}, \mathbf{b}, \mathbf{c}^T)$ as given in (6.37) is identically obtained with a structure-preserving finite volume discretization of the underlying PH model on staggered grids [95].

We have shown that for $\alpha = 0$, the output y_1 according to (6.36b) is flat, just as the original output of the distributed-parameter model (6.24) with the downstream boundary condition (6.26). A flat parametrization of the finite-dimensional state vector[9] $\mathbf{x} = \begin{bmatrix} \tilde{\mathbf{p}}^T & \tilde{\mathbf{q}} \end{bmatrix}^T = \begin{bmatrix} \mathbf{p}^T & \mathbf{q}^T \end{bmatrix}^T$ and the input u_1 based on the state space model (6.37) will have the form (6.3). In particular, the input $u_1^*(t)$ will depend on $y_1^*(t)$ *and* its first $2N$ time derivatives, which must exist and be bounded. This smoothness requirement on the desired output trajectory is at odds with the exact solution $u_1^*(t) = -y_1^*(t+1)$ determined in the previous subsection.

In order to get rid of the smoothness condition $y_1^*(t) \in C^{2N}([0, \infty))$, we perform a time discretization of the continuous-time approximate model with the goal to compute a *discrete-time feedforward control* sequence $u^{*,k}$ based on a desired output sequence $y^{*,k}$, $k \in \mathbb{N}_0$.

6.3.3 Full Discretization

We follow the lines of the previous chapter, in which discrete-time port-Hamiltonian systems based on *symplectic* time integration were introduced. We choose the *symplectic Euler* integration scheme, see [76], Section I.1.2, as the simplest structure-preserving integration method. In preparation for the nonlinear example in the next section, we do not substitute the boundary condition (6.26), but we keep the MIMO state representation $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ according

[9] Note that $\tilde{\mathbf{p}} = \mathbf{p}$, $\tilde{\mathbf{q}} = \mathbf{q}$ because of $\mathbf{P}_{fp} = \mathbf{P}_{fq} = \mathbf{I}$ in the case $\alpha = 0$.

to (6.35) with matrices

$$\mathbf{A} = \frac{1}{h}\begin{bmatrix} & & & & & -1 & & & \\ & & & & & 1 & -1 & & \\ & & & & & & \ddots & \ddots & \\ & & & & & & & 1 & -1 \\ 1 & -1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & 1 & -1 & & & & & \\ & & & 1 & & & & & \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 1 & \\ 0 & \\ \vdots & \\ 0 & \\ & 0 \\ & \vdots \\ & 0 \\ & -1 \end{bmatrix},$$
$$\mathbf{C} = \frac{1}{h}\begin{bmatrix} & & & & 0 & \dots & 0 & -1 \\ 1 & 0 & \dots 0 & & & & & \end{bmatrix}. \tag{6.39}$$

Discretizing the differential equations for $\mathbf{p}$ with the forward (explicit) Euler scheme, and the ones for $\mathbf{q}$ with backward (implicit) Euler, we obtain the system of *difference* equations (we now use Δt and Δz for the time and spatial step, respectively)

$$\begin{bmatrix} p_1^{k+1} - p_1^k \\ p_2^{k+1} - p_2^k \\ \vdots \\ p_N^{k+1} - p_N^k \\ q_1^{k+1} - q_1^k \\ \vdots \\ q_{N-1}^{k+1} - q_{N-1}^k \\ q_N^{k+1} - q_N^k \end{bmatrix} = \frac{\Delta t}{\Delta z}\begin{bmatrix} & & & & & -1 & & & \\ & & & & & 1 & -1 & & \\ & & & & & & \ddots & \ddots & \\ & & & & & & & 1 & -1 \\ 1 & -1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & 1 & -1 & & & & & \\ & & & 1 & & & & & \end{bmatrix}\begin{bmatrix} p_1^{k+1} \\ p_2^{k+1} \\ \vdots \\ p_N^{k+1} \\ q_1^k \\ \vdots \\ q_{N-1}^k \\ q_N^k \end{bmatrix} + \Delta t \begin{bmatrix} 1 & \\ 0 & \\ \vdots & \\ 0 & \\ & 0 \\ & \vdots \\ & 0 \\ & -1 \end{bmatrix}\begin{bmatrix} u_1^k \\ u_2^{k+1} \end{bmatrix},$$
$$\begin{bmatrix} y_1^k \\ y_2^k \end{bmatrix} = \frac{1}{\Delta z}\begin{bmatrix} -q_N^k \\ p_1^k \end{bmatrix}. \tag{6.40}$$

The boundary condition $u_2(t) = -y_1(t)$ injects damping to the system. In the continuous-time SISO model, this is represented by the negative semi-definite matrix $-\mathbf{B}_q\mathbf{B}_q^T$ in (6.36a). To be consistent with the continuous-time case, we set

$$u_2^{k+1} = -y_1^k, \tag{6.41}$$

which has the same effect to the discrete-time state matrix in (6.40).

Remark 6.4. If, as an alternative, the feedforward control problem for a given output y_2 shall be solved with the boundary condition $u_1 = -y_2$, the backward Euler scheme must be applied to the p differential equations and forward Euler to the q equations. Then, the discrete-time analogue of (6.41) is $u_1^{k+1} = -y_2^k$. (Without permuting the integration schemes, the discrete-time interconnection would be acausal.)

Remark 6.5. The i-th line of (6.40) can be rewritten in terms of only the p-variables as

$$p_i^{k+1} - 2p_i^k + p_i^{k-1} = \left(\frac{\Delta t}{\Delta z}\right)^2 (p_{i-1}^k - 2p_i^k + p_{i+1}^k), \tag{6.42}$$

which is the well-known conservative *leapfrog scheme*, see e. g. [88], Section 17.4.

The structure of the discrete-time finite-dimensional model (6.40) allows to solve the flatness-based feedforward control problem in discrete time. Inspection of the single equations yields the following algorithm.

Given downstream data:

$$y_1^k, \qquad u_2^{k+1} = -y_1^k \tag{6.43}$$

Initial step:

$$\begin{aligned} q_N^k &= -\Delta z y_1^k \\ p_N^{k+1} &= \Delta z u_2^{k+1} + \frac{\Delta z}{\Delta t}(q_N^{k+1} - q_N^k) \end{aligned} \tag{6.44}$$

For $i = N-1$ to 1 do:

$$\begin{aligned} q_i^k &= q_{i+1}^k + \frac{\Delta z}{\Delta t}(p_{i+1}^{k+1} - p_{i+1}^k) \\ p_i^{k+1} &= p_{i+1}^{k+1} + \frac{\Delta z}{\Delta t}(q_i^{k+1} - q_i^k) \end{aligned} \tag{6.45}$$

Resulting upstream input:

$$u_1^k = \frac{1}{\Delta z} q_1^k + \frac{1}{\Delta t}(p_1^{k+1} - p_1^k) \tag{6.46}$$

Theorem 6.4. The discrete-time finite-dimensional model (6.40), which together with (6.41) approximates the wave equation (6.24) under the algebraic boundary condition (6.26), has the flat output y_1^k. State and input trajectories can be computed explicitly in terms of the output sequence $\{y_1^k\}_{-K<k<K}$ on a sufficiently large discrete time window $\{-K\Delta t, \ldots, 0, \ldots, K\Delta t\}$. Bounded sequences $\{p_i^k\}_{-K<k<K}$ and $\{q_i^k\}_{-K<k<K}$, $i = 1, \ldots, N$, for the states and a bounded feedforward control sequence $\{u_1^k\}_{-K<k<K}$ for arbitrary N are obtained under the condition $\frac{\Delta z}{\Delta t} \leq 1$.

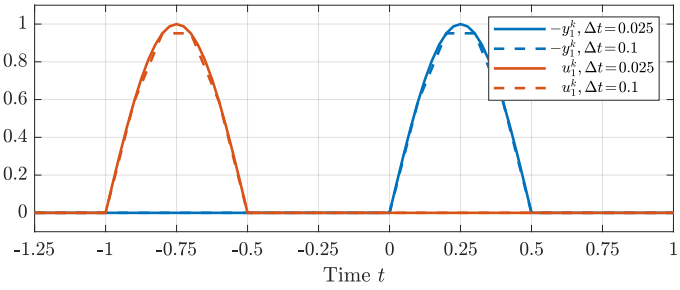

Figure 6.3: Desired output $-y_1^k$ (sine half wave) and computed input u_1^k for two different sampling times. Grid size: $\Delta z = \Delta t$.

Proof. Flatness of the output y_1^k follows directly from the algorithm (6.43)–(6.46), in which each step only depends on data from the previous one. The condition for boundedness of the computed state and input sequences follows from inspection of the algorithm, which involves the multiplication with $\frac{\Delta z}{\Delta t}$ in every step. To compute u_1^k, for example, the delayed/advanced flat output y_1^k is scaled with $\left(\frac{\Delta z}{\Delta t}\right)^{2N}$, which remains bounded for $N \to \infty$ only if $\Delta z \leq \Delta t$. □

Remark 6.6. The condition $\Delta z \leq \Delta t$ for the considered wave equation with propagation speed $c = 1$ is an *inverse* CFL condition. This is plausible in view of the considered feedforward control problem, which is an inverse boundary value problem.

Algorithm in compact notation

We can collect the system quantities over discrete time intervals of length $2K+1$ in the vectors for the states

$$\mathbf{P}_i = \begin{bmatrix} p_i^{-K+1} \\ \vdots \\ p_i^1 \\ \vdots \\ p_i^{K+1} \end{bmatrix}, \quad \mathbf{Q}_i = \begin{bmatrix} q_i^{-K} \\ \vdots \\ q_i^0 \\ \vdots \\ q_i^K \end{bmatrix}, \qquad i = 1, \ldots, N, \tag{6.47}$$

the given downstream data and the unknown upstream input:

$$\mathbf{Y}_1 = \begin{bmatrix} y_1^{-K} \\ \vdots \\ y_1^0 \\ \vdots \\ y_1^K \end{bmatrix}, \quad \mathbf{U}_2 = \begin{bmatrix} u_2^{-K+1} \\ \vdots \\ u_2^1 \\ \vdots \\ u_2^{K+1} \end{bmatrix}, \qquad \mathbf{U}_1 = \begin{bmatrix} u_1^{-K} \\ \vdots \\ u_1^0 \\ \vdots \\ u_1^k \end{bmatrix}. \tag{6.48}$$

With $\mathbf{S}_1 = \mathrm{diag}_1\{1,\dots,1\}$ and $\mathbf{S}_{-1} = \mathrm{diag}_{-1}\{1,\dots,1\}$ the upper and lower *shift matrices*, the algorithm (6.41)–(6.46) reads as follows.

Given downstream data:

$$\mathbf{Y}_1, \quad \mathbf{U}_2 = -\mathbf{Y}_1 \tag{6.49}$$

Initial step:

$$\begin{aligned} \mathbf{Q}_N &= -\Delta z \mathbf{Y}_1 \\ \mathbf{P}_N &= \Delta z \mathbf{U}_2 + \frac{\Delta z}{\Delta t}(\mathbf{S}_1 - \mathbf{I})\mathbf{Q}_N \end{aligned} \tag{6.50}$$

For $i = N-1$ **to** 1 **do:**

$$\begin{aligned} \mathbf{Q}_i &= \mathbf{Q}_{i+1} + \frac{\Delta z}{\Delta t}(\mathbf{I} - \mathbf{S}_{-1})\mathbf{P}_{i+1} \\ \mathbf{P}_i &= \mathbf{P}_{i+1} + \frac{\Delta z}{\Delta t}(\mathbf{S}_1 - \mathbf{I})\mathbf{Q}_i \end{aligned} \tag{6.51}$$

Resulting upstream input:

$$\mathbf{U}_1 = \frac{1}{\Delta z}\mathbf{Q}_1 + \frac{1}{\Delta t}(\mathbf{I} - \mathbf{S}_{-1})\mathbf{P}_1 \tag{6.52}$$

Note that the parameter K must be chosen large enough such that no information is lost when the vectors $\mathbf{Q}_i$ and $\mathbf{P}_{i+1}$ are pre-multiplied with the shift matrices $\mathbf{S}_1$ and $\mathbf{S}_{-1}$, respectively.

Figure 6.3 show the result of applying this scheme with a sine half wave as the desired output $y^*(t)$ starting at $t = 0$. The exact solution of the feedforward control problem according to (6.34) is a prediction of this sine half wave, $u^*(t) = -y^*(t+1)$. For the chosen step sizes – the length 1 of the spatial domain and the delay 1 are integer multiples of them – the algorithm gives a perfectly shifted desired output signal as the control input. Under grid refinement, the signal shape converges to the continuous-time sine half wave.

6.4 Nonlinear Hyperbolic Systems

Inspired by the result of the previous section, we consider a one-dimensional hyperbolic system of two conservation laws on $\Omega = (0, L)$ with *nonlinear* constitutive equations of the form

$$e^p = \frac{1}{2}(*q)^2 + F(*p, z), \qquad e^q = *p*q, \tag{6.53}$$

where F is an arbitrary function, which is invertible with respect to $*p$ in the considered operating domain. Note that both the 1D version of the shallow water equations treated below, and the Euler equations of isentropic gas flow, see e. g. [137], feature constitutive equations in this form.

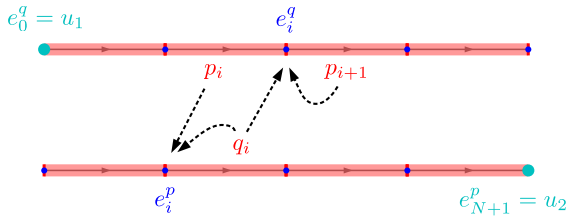

Figure 6.4: Spatial dependencies in the discretized constitutive equations.

6.4.1 Consistently Discretized Constitutive Equations

Applying a spatial discretization of the structure equations (6.24a) with the simplest Whitney forms and the parameter $\alpha = 0$, the lumped states $\tilde{p}_i = p_i$ and $\tilde{q}_i = q_i$, $i = 1, \ldots, N$, have the interpretation of *integral* conserved quantities over the discretization intervals (edges). Consequently, $\frac{p_i}{\Delta z}$ and $\frac{q_i}{\Delta z}$ with $\Delta z = \frac{L}{N}$ denote the average values on each interval. On the other hand, the degrees of freedom e_i^p and e_i^q, $i = 1, \ldots, N$, approximate the values of the co-states in the discretization nodes, where no boundary conditions are imposed. A consistent approximation of the constitutive equations (6.53) in the discretization nodes based on the states on adjacent edges is given by

$$e_i^p = \frac{1}{2}\left(\frac{q_i}{\Delta z}\right)^2 + F_i\left(\frac{p_i}{\Delta z}\right), \qquad e_i^q = \frac{p_{i+1} q_i}{\Delta z^2}. \tag{6.54}$$

$F_i(\cdot) := F(\cdot, z_i)$ denotes the evaluation of the function F in the nodal coordinates $z_i^p = (i-1)\Delta z$. The particular choice (6.54) guarantees a flat parametrization of the state and the input u_1 in terms of a given desired output y_1, see further below. The dependencies of the discrete efforts on the neighboring states are illustrated in Fig. 6.4.

Remark 6.7. The proof that (6.54) consistently approximates the local version of the constitutive equations (6.53) can be performed along the same lines as for the 2D case sketched in Appendix B.1. Note, that e_i^p – in contrast to e_i^q – is computed based on the downstream approximations of *both* states.

6.4.2 Discrete-Time State Representation

The discrete-time state representation that results from application of the symplectic Euler scheme is, in analogy to (6.40),

$$\begin{bmatrix} p_1^{k+1}-p_1^k \\ p_2^{k+1}-p_2^k \\ \vdots \\ p_N^{k+1}-p_N^k \\ q_1^{k+1}-q_1^k \\ \vdots \\ q_{N-1}^{k+1}-q_{N-1}^k \\ q_N^{k+1}-q_N^k \end{bmatrix} = \Delta t \begin{bmatrix} & & & & & & -1 & & \\ & & & & & & 1 & -1 & \\ & & & & & & & \ddots & \ddots \\ & & & & & & & & 1 & -1 \\ 1 & -1 & & & & & & & & \\ & \ddots & \ddots & & & & & & & \\ & & 1 & -1 & & & & & & \\ & & & 1 & & & & & & \end{bmatrix} \begin{bmatrix} e_1^{p,k+1} \\ e_2^{p,k+1} \\ \vdots \\ e_N^{p,k+1} \\ e_1^{q,k} \\ \vdots \\ e_{N-1}^{q,k} \\ e_N^{q,k} \end{bmatrix} + \Delta t \begin{bmatrix} 1 & \\ 0 & \\ \vdots & \\ 0 & \\ & 0 \\ & \vdots \\ & 0 \\ & -1 \end{bmatrix} \begin{bmatrix} u_1^k \\ u_2^{k+1} \end{bmatrix},$$

$$\begin{bmatrix} y_1^k \\ y_2^k \end{bmatrix} = \begin{bmatrix} -e_N^{q,k} \\ e_1^{p,k} \end{bmatrix}. \tag{6.55}$$

Taking into account that the symplectic Euler scheme requires the evaluation of right hand side quantities at discrete times $k+1$ and k, respectively, we write the discrete-time version of the constitutive equations (6.54) as

$$e_i^{p,k+1} = \frac{1}{2}\left(\frac{q_i^{k+1}}{\Delta z}\right)^2 + F_i\left(\frac{p_i^{k+1}}{\Delta z}\right), \qquad e_i^{q,k} = \frac{p_{i+1}^k q_i^k}{\Delta z^2}. \tag{6.56}$$

6.4.3 Flatness-Based Feedforward Control

We adapt the algorithm (6.43)–(6.46) to the nonlinear case with a downstream boundary condition

$$B(u_2^k, y_1^k) = 0. \tag{6.57}$$

From this discrete version of (6.25), the collocated input u_2^k can be computed for given y_1^k. Taking into account (6.55) and (6.56), the following iteration produces the state and effort variables on the spatio-temporal discretization grid. Under the indicated assumptions, p_{N+1}^k can be computed in an initial step from the algebraic equation, which results from substitution of (6.61a) in (6.61c). This auxiliary state allows to start the iteration loop with $i = N$.

Given downstream data:

$$e_N^{q,k} = -y_1^k, \qquad e_{N+1}^{p,k} = u_2^k \tag{6.58}$$

Assumptions:

$$q_{N+1}^k = q_N^k, \qquad F_{N+1}(\cdot) = F_N(\cdot) \tag{6.59}$$

Initial step:

$$\text{Compute} \quad p_{N+1}^k \quad \text{from} \qquad F_N\left(\frac{p_{N+1}^k}{\Delta z}\right) = u_2^k - \frac{1}{2}\left(\frac{y_1^k \Delta z}{p_{N+1}^k}\right)^2 \tag{6.60}$$

For $i = N$ to 1 do:

$$q_i^k = \Delta z^2 \frac{e_i^{q,k}}{p_{i+1}^k} \tag{6.61a}$$

$$e_i^{p,k+1} = e_{i+1}^{p,k+1} + \frac{1}{\Delta t}(q_i^{k+1} - q_i^k) \tag{6.61b}$$

$$p_i^{k+1} = \Delta z F_i^{-1}\left(e_i^{p,k+1} - \frac{1}{2}\left(\frac{q_i^{k+1}}{\Delta z}\right)^2\right) \tag{6.61c}$$

$$e_{i-1}^{q,k} = e_i^{q,k} + \frac{1}{\Delta t}(p_i^{k+1} - p_i^k) \tag{6.61d}$$

Resulting upstream input:

$$u_1^k = e_0^{q,k} \tag{6.62}$$

6.4.4 Example: 1D Shallow Water Equations

As an example for a nonlinear hyperbolic system of conservation laws with constitutive equations of the form (6.53) we consider the one-dimensional shallow water or Saint-Venant equations. The inverse or flow routing problem for this system was solved in [181] using the implicit box scheme. The flatness-based generation of input trajectories was addressed in [93] using the method of characteristics. Flatness-based control based on a parabolic approximation, the Hayami model, is presented in [159]. In [99], the feedforward control problem was solved based on the port-Hamiltonian model and its structure-preserving discretization according to [71]. The feedthrough of the finite-dimensional model allowed for a direct model inversion and an iterative computation of the input trajectory with the method presented in [49].

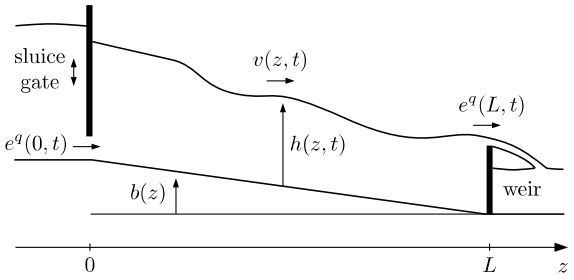

Figure 6.5: Sketch of a channel with descending slope.

Model and algorithm

On a one-dimensional domain $\Omega = (0, L)$, the structured model (2.73) for the two-dimensional shallow water equations with Hamiltonian functional (2.75) boils down[10] to

$$\dot{p} = -\mathrm{d}e^q \tag{6.63a}$$

$$\dot{q} = -\mathrm{d}e^p - r(*q)*q. \tag{6.63b}$$

$p = h(z)dz$ and $q = v(z)dz$ are the one-forms related to the water depth $h(z)$ and the flow velocity $v(z)$, respectively. The additional dissipation term with velocity-dependent coefficient

$$r(*q) = \frac{gC|*q|^{m-1}}{R^P} \tag{6.64}$$

takes into account friction losses. R denotes the hydraulic radius, which depends on the channel cross section, P depends on the friction model and C is a coefficient. The parameter m characterizes the flow type[11]. The 1D constitutive equations (the invertible function F according to (6.54) is affine here) read

$$e^p = \frac{1}{2}(*q)^2 + g(*p + b), \qquad e^q = *p*q, \tag{6.65}$$

where e^p represents the *total head* scaled with g (unit $\mathrm{m}^2/\mathrm{s}^2$), and e^q the *specific discharge* (unit m^2/s). The function $b : [0, L] \to \mathbb{R}$ describes the bottom profile of the channel. The downstream discharge relation (flow over a sharp-crested weir of height w) is given by[12]

$$e^q(L) = C_w \frac{2}{3}\sqrt{2g}(*p(L) - w)^{\frac{3}{2}} \tag{6.66}$$

with a constant discharge coefficient C_w.

[10] The vorticity term vanishes in 1D.

[11] See e. g. [38], Section 12-2: $m = 1$ represents laminar flow, $m = 1.75$ smooth, turbulent flow, and $m = 2$ fully rough, turbulent flow.

[12] See [38], Section 7-5.

The general algorithm for discrete-time flatness-based feedforward control from the previous section can be adapted as follows to this particular case. Besides the nonlinear boundary conditions, the distributed and velocity-dependent friction has to be considered.

Given downstream data:

$$h^k(L), \qquad p_{N+1}^k = \Delta z h^k(L) \tag{6.67}$$

Assumption:

$$q_{N+1}^k = q_N^k \tag{6.68}$$

Downstream variables:

$$-y_1^k = e_N^{q,k} = C_w \frac{2}{3}\sqrt{2g}\left(\frac{p_{N+1}^k}{\Delta z} - w\right)^{\frac{3}{2}} \tag{6.69a}$$

$$q_{N+1}^k = q_N^k = \Delta z^2 \frac{e_N^{q,k}}{p_{N+1}^k} \tag{6.69b}$$

$$u_2^{k+1} = e_{N+1}^{p,k+1} = \frac{1}{2}\left(\frac{q_{N+1}^{k+1}}{\Delta z}\right)^2 + g\frac{p_{N+1}^{k+1}}{\Delta z} \tag{6.69c}$$

For $i = N$ to 1 do:

$$q_i^k = \Delta z^2 \frac{e_i^{q,k}}{p_{i+1}^k} \qquad \text{(skip for } i = N\text{)} \tag{6.70a}$$

$$e_i^{p,k+1} = e_{i+1}^{p,k+1} + \frac{1}{\Delta t}(q_i^{k+1} - q_i^k) + r\left(\frac{q_i^k}{\Delta z}\right) q_i^k \tag{6.70b}$$

$$p_i^{k+1} = \Delta z \left(\frac{1}{g}\left(e_i^{p,k+1} - \frac{1}{2}\left(\frac{q_i^{k+1}}{\Delta z}\right)^2\right) - b((i-1)\Delta z)\right) \tag{6.70c}$$

$$e_{i-1}^{q,k} = e_i^{q,k} + \frac{1}{\Delta t}(p_i^{k+1} - p_i^k) \tag{6.70d}$$

Resulting upstream input:

$$u_1^k = e_0^{q,k} \tag{6.71}$$

In order to start the algorithm, we associate the water depth at the weir with the auxiliary variable p_{N+1}^k. Moreover, we assume $q_{N+1}^k = q_N^k$, which is justified by the continuity of the flow velocity around the node representing the sharp-crested weir.

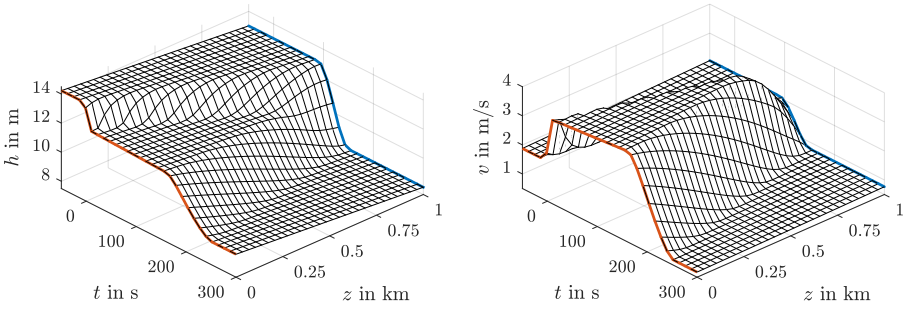

Figure 6.6: Computed profiles of flow depth and velocity over distance and time.

Numerical results

We validate our algorithm with the scenario and channel parameters from [93]. A transient of the downstream water depth in front of an overflow weir shall be realized using the upstream (specific) discharge, which can be regulated by the opening of an upstream sluice gate. The channel parameters, as well as the computation of the steady state solutions (without friction) and the discussion of the flow regime can be found in Appendix B.3.

Figure 6.6 shows the surface plots of depth and velocity over the spatial variable and time. The level sets of both state variables in the (z,t) plane before and after the transient nicely illustrate the characteristics of the hyperbolic system.

Figure 6.7 depicts up- and downstream water depths and velocities for a desired downstream depth $h^k(L)$ under different step sizes Δz with $\Delta t = 2\Delta z$. The discrete states p_i^k and q_i^k correspond to integral values of depth and velocity on the discretization intervals, therefore their mean values $p_i^k/\Delta z$ and $q_i^k/\Delta z$ can be localized in the interval centers $(i-\frac{1}{2})\Delta z$. To obtain numerical values for depth and velocity at $z=0$ and $z=L$, we compute

$$h^k(0) = \Delta z \frac{e_1^{q,k}}{q_1^k}, \quad v^k(0) = \Delta z \frac{e_1^{q,k}}{p_1^k}, \qquad v^k(L) = \Delta z \frac{e_N^{q,k}}{p_N^k}. \tag{6.72}$$

With decreasing step size Δz, the numerical solutions converge, and match the results depicted in [93], Fig. 7, which have been obtained using the method of characteristics.

Remark 6.8. Concerning temporal and spatial step size and their ratio, we note the following. In the example, the minimum characteristic speed[13] is around $c_{low} = 7.5\,\mathrm{m/s}$, such that $\Delta t = k\Delta z$ with $k > \frac{1}{c_{low}} \approx 0.13\,\mathrm{s/m}$ must be chosen to satisfy the inverse CFL condition. The depicted trajectory with

[13]The state-dependent characteristic velocities $v \pm \sqrt{gh}$ for the shallow water equations emerge after transformation to Riemann coordinates (invariants), see e. g. [45], Section 7.3.

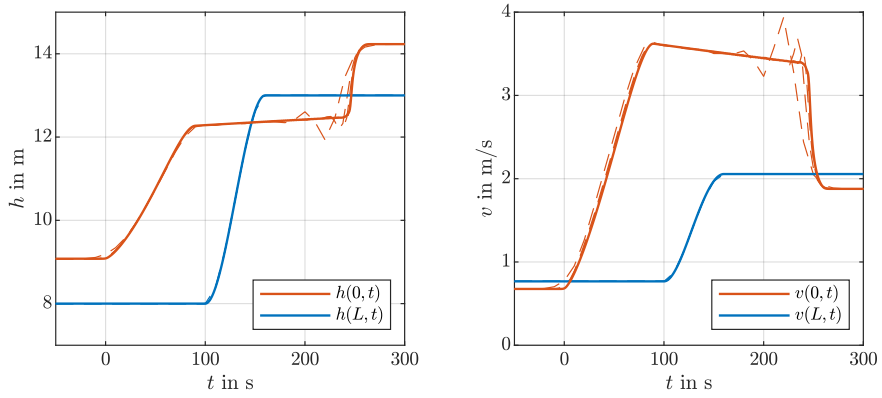

Figure 6.7: Downstream (given) and upstream (computed) depth and velocity over time. The plots illustrate the convergence of solutions with decreasing step size $\Delta z \in \{5, 2, 1, 0.5, 0.25\}$ and $\Delta t = 2\Delta z$.

steep upstream transients requires high values of k, e. g. (skipping units) $k = 2$, for satisfactory convergence. For slower transients, e. g. when the transition time is extended by 100 s, k can be chosen smaller, with at the same time a coarser spatial grid. In any case, choosing an extremely fine spatial grid provokes numerical oscillations which motivates the choice of an *optimal* grid size, adapted to the system and the desired trajectory.

6.5 Conclusions

The semi-discretization scheme from Chapter 4 does not only preserve the port-Hamiltonian structure of the open system model. Using the simplest Whitney forms in one spatial dimension, also the flatness of a given boundary output is conserved under an appropriate choice of the discretization parameter α. For the parabolic heat equation, the input trajectory generation using a finite number of time derivatives of the flat output is consistent with the infinite-dimensional solution of the inverse problem. In the hyperbolic case, additional geometric time integration leads to discrete-time finite-dimensional models, which correspond to the well-known *leapfrog* scheme, and based on which feedforward controllers can be computed without unnecessary smoothness constraints on the output. With a consistent approximation of the constitutive equations, also the solution to the flow routing problem for the Saint-Venant equations can be computed fast, efficiently and without integration.

The presented study is the basis for considering flatness conservation in discretized PH models and the numerical trajectory generation (i) in more than one spatial dimension (see [133] for a spectral approach) and (ii) with non-local approximation bases (like Lagrange interpolation polynomials, see [138]) in the spatial discretization step.

Appendix A

Mathematical Background

A.1 Exterior Differential Calculus

We give a compact introduction to the calculus with differential forms and their functional spaces. For further reading, we refer to [59], [8] and the paper [6] with its numerous references. The calculus with differential forms or *exterior differential calculus* is for example used for the representation and numerical simulation of Maxwell's equations [21], to give one example. *Discrete exterior calculus* [48] extends the formalism to discrete geometric objects defined on oriented meshes, and *finite element exterior calculus* [6] sets the framework for numerical approximation using finite element spaces of differential forms [7].

An essential characterization of (exterior) differential forms (of degree k, or k-forms) is given on page 1 of [59] as "*things which occur under integral signs*". However, they are not merely "densities", but they have an *orientation*, i. e. they contain the information about the sense of integration.

A.1.1 Smooth Differential Forms

We represent distributed parameter PH systems in the language of *differential forms*, see e. g. [59] for a comprehensive introduction to *smooth* differential forms, i. e. differential forms with sufficiently differentiable (in the classical sense) coefficient functions. Let Ω be an open, bounded and connected n-dimensional spatial domain with Lipschitz boundary $\partial\Omega$ and denote $\Lambda^k(\Omega)$ the space of smooth differential k-forms on Ω. For a smooth $(n-1)$-form $\omega \in \Lambda^{n-1}(\Omega)$, the continuous extension to the boundary is denoted $\operatorname{tr}\omega \in \Lambda^{n-1}(\partial\Omega)$. The symbol tr stems from the *trace map*, which defines the extension to the boundary for Lebesgue integrable functions (see further below). The *wedge product* $\wedge : \Lambda^k(\Omega) \times \Lambda^l(\Omega) \to \Lambda^{k+l}(\Omega)$ is a skew-symmetric exterior product of differential forms. The *exterior derivative* $\mathrm{d} : \Lambda^k(\Omega) \to \Lambda^{k+1}(\Omega)$ with $\mathrm{d} \circ \mathrm{d} = 0$, is a unique differential operator on differential forms of degree k. The sequence of spaces of differential forms, connected via the exterior

derivative, is the so-called *de Rham complex*[1].

Typical examples from electromagnetism[2] in $\mathbb{R}^3$ are the electric field one-form $E_x dx + E_y dy + E_z dz$, the current density 2-form $J_x dy \wedge dz + J_y dz \wedge dx + J_z dx \wedge dy$ or the charge density 3-form $\rho\, dx \wedge dy \wedge dz$. In $\mathbb{R}^n$, the differentials $\{dx_1, \ldots, dx_n\}$ form the basis of differential one-forms. Higher order basis forms are constructed using the wedge (or exterior) product, which due to its skew-symmetry (see below) induces the orientation.

We will make frequent use of the following formulas[3] for $\lambda \in \Lambda^k(\Omega)$ and $\mu \in \Lambda^l(\Omega)$, which express the skew-symmetry of the wedge product and the product rule for the exterior derivative:

$$\lambda \wedge \mu = (-1)^{kl} \mu \wedge \lambda, \tag{A.1}$$

$$\mathrm{d}(\lambda \wedge \mu) = \mathrm{d}\lambda \wedge \mu + (-1)^k \lambda \wedge \mathrm{d}\mu. \tag{A.2}$$

A natural pairing or *duality product* between two differential forms $\lambda \in \Lambda^k(\Omega)$ and $\mu \in \Lambda^{n-k}(\Omega)$ on Ω is given by

$$\langle \lambda | \mu \rangle_\Omega := \int_\Omega \lambda \wedge \mu. \tag{A.3}$$

The duality product is defined accordingly on the $(n-1)$-dimensional boundary $\partial\Omega$ of Ω, see [197], Eq. (5).

The *Hodge star* induces an *inner product* on the space of differential forms on a manifold Ω by

$$(\alpha, \beta) := \langle \alpha | * \beta \rangle_\Omega = \langle \beta | * \alpha \rangle_\Omega = (\beta, \alpha), \qquad \alpha, \beta \in \Lambda^k(\Omega), \tag{A.4}$$

see Section 8.4 of [59] or Section 3.6 of [86]. The inner product is not necessarily the standard L^2 inner product (A.8), but may be equipped with another metric, see e. g. the energy norm for linear PH systems [89]. The Hodge star is, hence, *metric* dependent. A given inner product space induces a corresponding Hodge star. The subsequent application of the Hodge star may change the sign of the original differential form: $**\alpha = (-1)^{k(n-k)}\alpha$.

Index raising ($\sharp$) produces a vector field with the same components from a one-form. Index lowering ($\flat$) produces a one-form with identical components from a vector field. Raising and lowering in these *musical isomorphisms* refers to the fact that upper (lower) indices are typically used for the components of vector fields (one-forms).

With the Hodge star on $\mathbb{R}^3$ and the musical isomorphisms, the differential operators from vector calculus can be expressed in terms of the exterior derivative:

$$\operatorname{grad} f = (\mathrm{d}f)^\sharp, \qquad \operatorname{rot} \mathbf{g} = (*(\mathrm{d}\mathbf{g}^\flat))^\sharp, \qquad \operatorname{div} \mathbf{g} = *\mathrm{d}(*\mathbf{g}^\flat). \tag{A.5}$$

[1] The complex property $\mathrm{d} \circ \mathrm{d} = 0$ is well-known from vector calculus, where the subsequent application of differential operators maps to zero: $\operatorname{rot} \operatorname{grad} f = 0$, $\operatorname{div} \operatorname{rot} \mathbf{g} = \mathbf{0}$.

[2] See e. g. the recent article [208] for an illustrative introduction to Maxwell's equations in terms of differential forms, with an abundant list of references.

[3] See e. g. [59], Sections 2.3 and 3.2.

A.1.2 Stokes' Theorem

Using exterior differential calculus, the *generalized Stokes' theorem*[4] unifies the different integration formulas from vector calculus:

Theorem A.1. Let ω be a differential $(n-1)$-form on Ω. Then

$$\int_\Omega \mathrm{d}\omega = \int_{\partial\Omega} \operatorname{tr}\omega. \tag{A.6}$$

The generalized Stokes' theorem (A.1), together with the product rule (A.2) and the short notation of the duality product (A.3), gives the *integration-by-parts* formula for smooth differential forms $\lambda \in \Lambda^k(\Omega)$ and $\mu \in \Lambda^{n-k-1}(\Omega)$,

$$\langle \mathrm{d}\lambda | \mu \rangle_\Omega = \langle \operatorname{tr}\lambda | \operatorname{tr}\mu \rangle_{\partial\Omega} - (-1)^k \langle \lambda | \mathrm{d}\mu \rangle_\Omega. \tag{A.7}$$

A.1.3 Lebesgue and Sobolev Spaces of Differential Forms

We recall some important definitions and facts, which ensure that the formulas from the previous subsection make also sense on functional spaces of differential forms with weaker smoothness conditions. Section 4 of [6] gives a quick and concise introduction into calculus with differential forms whose coefficient functions belong to Lebesgue spaces $L^p(\Omega)$ and Sobolev spaces, in particular $H^m(\Omega) = W^{m,2}(\Omega)$. The space $L^2\Lambda^k(\Omega)$ of differential forms with square integrable coefficient functions is equipped with the inner product[5]

$$(\alpha, \beta)_{L^2\Lambda^k(\Omega)} := \int_\Omega \sum_{i=1}^N \alpha_i(z)\beta_i(z)\, d\mathrm{vol}, \tag{A.8}$$

where $\alpha_i, \beta_i \in L^2(\Omega)$, $i = 1, \ldots, N$, are the coefficient functions of $\alpha, \beta \in L^2\Lambda^k(\Omega)$. The *weak exterior derivative* $\mathrm{d}\lambda$ of $\lambda \in \Lambda^k(\Omega)$ can be defined via the integration-by-parts formula (A.7), with smooth differential forms μ that vanish on the boundary (due to their compact support in Ω):

$$\langle \mathrm{d}\lambda | \mu \rangle_\Omega = -(-1)^k \langle \lambda | \mathrm{d}\mu \rangle_\Omega \qquad \forall \mu \in C_c^\infty \Lambda^{n-k-1}(\Omega). \tag{A.9}$$

We do not introduce a new symbol, as we will understand d in this weak sense in the whole work. This allows to apply the exterior derivative to differential forms whose coefficient functions are not differentiable in the classical sense. The Sobolev spaces $H^m\Lambda^k(\Omega)$ contain the differential forms on Ω with L^2

[4] See e. g. [59], Section 5.8 or [8], Section 36.D, where it is expressed for Ω a n-chain, i. e. a formal sum of n-simplices on a manifold $M \supset \Omega$ and nicely called *Newton-Leibniz-Gauss-Green-Ostrogradskii-Stokes-Poincaré formula*.

[5] To define the inner product, we need a volume form. For $\Omega \subset \mathbb{R}^n$, we take $d\mathrm{vol} = d^n z$ as in [86], Definition 3.6.2.

weak derivatives up to order m. The corresponding inner product for $m = 1$ is defined as

$$(\alpha, \beta)_{H^1\Lambda^k(\Omega)} := (\alpha, \beta)_{L^2\Lambda^k(\Omega)} + (\mathrm{d}\alpha, \mathrm{d}\beta)_{L^2\Lambda^{k+1}(\Omega)}. \tag{A.10}$$

As we deal with *boundary control systems*, we are particularly interested in the extension of certain differential forms to the boundary. Fortunately, the *trace theorem* from classical functional analysis[6] extends to differential forms as discussed in Section 4 of [6]. We will make heavy use of the implication

$$\lambda \in H^1\Lambda^k(\Omega) \quad \Rightarrow \quad \operatorname{tr}\lambda \ \in \ H^{1/2}\Lambda^k(\partial\Omega) \ \subset \ L^2\Lambda^k(\partial\Omega). \tag{A.11}$$

Where convenient for compactness, we use the common abusive notation $\int_{\partial\Omega} \omega = \int_{\partial\Omega} \operatorname{tr}\omega$ for the extension of $\omega \in H^m\Lambda^{n-1}(\Omega)$, $m \geq 1$ to the boundary.

A.2 Geometric Numerical Integration

The following Butcher tables show the nodes $c_1, \ldots, c_s$, the weights $b_1, \ldots, b_s$, and the Runge-Kutta coefficients a_{ij}, $i, j = 1, \ldots, s$, of the integration schemes used in Chapter 5. The tables can be found, for example, in [76], Sections II.1.3 and II.1.4.

Table A.1: Runge-Kutta coefficients for Gauss-Legendre methods, $s = 1, 2, 3$.

$s = 1$:

$\frac{1}{2}$	$\frac{1}{2}$
	1

$s = 2$:

$\frac{1}{2} - \frac{\sqrt{3}}{6}$	$\frac{1}{4}$	$\frac{1}{4} - \frac{\sqrt{3}}{6}$
$\frac{1}{2} + \frac{\sqrt{3}}{6}$	$\frac{1}{4} + \frac{\sqrt{3}}{6}$	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{2}$

$s = 3$:

$\frac{1}{2} - \frac{\sqrt{15}}{10}$	$\frac{5}{36}$	$\frac{2}{9} - \frac{\sqrt{15}}{15}$	$\frac{5}{36} - \frac{\sqrt{15}}{30}$
$\frac{1}{2}$	$\frac{5}{36} + \frac{\sqrt{15}}{24}$	$\frac{2}{9}$	$\frac{5}{36} - \frac{\sqrt{15}}{24}$
$\frac{1}{2} + \frac{\sqrt{15}}{10}$	$\frac{5}{36} + \frac{\sqrt{15}}{30} + \frac{\sqrt{3}}{6}$	$\frac{2}{9} + \frac{\sqrt{15}}{15}$	$\frac{5}{36}$
	$\frac{5}{18}$	$\frac{4}{9}$	$\frac{5}{18}$

[6] See e. g. [26], Section 9.8, pp. 315-316, or [158], Section 1.3, p. 10, on the introduction of the trace operator in terms of functional analysis.

Table A.2: Runge-Kutta coefficients of the 3- and 4-stage Lobatto IIIA methods.

$s = 3$:

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{5}{24} & \frac{1}{3} & -\frac{1}{24} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

$s = 4$:

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \frac{5-\sqrt{5}}{10} & \frac{11+\sqrt{5}}{120} & \frac{25-\sqrt{5}}{120} & \frac{25-13\sqrt{5}}{120} & \frac{-1+\sqrt{5}}{120} \\ \frac{5+\sqrt{5}}{10} & \frac{11-\sqrt{5}}{120} & \frac{25+13\sqrt{5}}{120} & \frac{25+\sqrt{5}}{120} & \frac{-1-\sqrt{5}}{120} \\ 1 & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \\ \hline & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{array}$$

Table A.3: Runge-Kutta coefficients of the 3- and 4-stage Lobatto IIIB methods.

$s = 3$:

$$\begin{array}{c|ccc} 0 & \frac{1}{6} & -\frac{1}{6} & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} & 0 \\ 1 & \frac{1}{6} & \frac{5}{6} & 0 \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

$s = 4$:

$$\begin{array}{c|cccc} 0 & \frac{1}{12} & \frac{-1-\sqrt{5}}{24} & \frac{-1+\sqrt{5}}{24} & 0 \\ \frac{5-\sqrt{5}}{10} & \frac{1}{12} & \frac{25+\sqrt{5}}{120} & \frac{25-13\sqrt{5}}{120} & 0 \\ \frac{5+\sqrt{5}}{10} & \frac{1}{12} & \frac{25+13\sqrt{5}}{120} & \frac{25-\sqrt{5}}{120} & 0 \\ 1 & \frac{1}{12} & \frac{11-\sqrt{5}}{24} & \frac{11+\sqrt{5}}{24} & 0 \\ \hline & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{array}$$

Appendix B

Computations

B.1 Consistency of the Finite Volume Approximation

This section contains the computations to prove the statements on the consistency order of the finite volume approximation of the 2D shallow water equations (SWE) on a 2D rectangular grid in Chapter 3.

B.1.1 Model in Terms of Average States

We sketch the transition from the integral to the average model. The matrices $\mathbf{J}$ and $\mathbf{G}$ in (3.28), (3.40) for the 2D SWE are partitioned as follows:

$$\mathbf{J} = \begin{bmatrix} \mathbf{0} & \mathbf{J}_1 & \mathbf{J}_2 \\ -\mathbf{J}_1^T & \mathbf{0} & \mathbf{0} \\ -\mathbf{J}_2^T & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{0} & \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{G}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{G}_4 & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{B.1}$$

Integral and average states and efforts are related via

$$\mathbf{x}_d = \boldsymbol{\Delta}^x \bar{\mathbf{x}}, \qquad \mathbf{e}_d = \boldsymbol{\Delta}^e \bar{\mathbf{e}}, \tag{B.2}$$

where

$$\boldsymbol{\Delta}^x = \begin{bmatrix} \Delta x \Delta y \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Delta x \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Delta y \mathbf{I} \end{bmatrix}, \qquad \boldsymbol{\Delta}^e = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Delta y \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Delta x \mathbf{I} \end{bmatrix}. \tag{B.3}$$

Note that $\boldsymbol{\Delta}^x \boldsymbol{\Delta}^e = \Delta x \Delta y \mathbf{I}$. With the discrete Hamiltonian density

$$\bar{H}_d(\bar{\mathbf{x}}) := \frac{1}{\Delta x \Delta y} H_d(\boldsymbol{\Delta}^x \bar{\mathbf{x}}) \tag{B.4}$$

such that

$$\begin{aligned} \nabla \bar{H}_d(\bar{\mathbf{x}}) &= \frac{1}{\Delta x \Delta y} \boldsymbol{\Delta}^x \nabla H_d(\mathbf{x}_d) \\ &= (\boldsymbol{\Delta}^e)^{-1} \nabla H_d(\mathbf{x}_d), \end{aligned} \tag{B.5}$$

(3.40a) transforms to the average state differential equation (3.41) with $\bar{\mathbf{e}} = \nabla \bar{H}_d(\bar{\mathbf{x}})$ and

$$\bar{\mathbf{J}} = \Delta x (\mathbf{\Delta}^x)^{-1} \mathbf{J} \mathbf{\Delta}^e, \qquad \bar{\mathbf{G}} = \Delta x (\mathbf{\Delta}^x)^{-1} \mathbf{G} \mathbf{\Delta}^e. \tag{B.6}$$

For $\Delta x = \Delta y$, we obtain $\bar{\mathbf{J}} = \mathbf{J}$ and $\bar{\mathbf{G}} = \mathbf{G}$.

B.1.2 Computations of Local Errors

For the Hamiltonian density (3.30) with $b \equiv 0$, the average discrete efforts $\bar{\mathbf{e}}$ according to (3.39), have the components

$$\begin{aligned} \bar{e}_j^h &= \frac{1}{2} \left(\frac{u_{d,\mathrm{le}}^2 + u_{d,\mathrm{ri}}^2}{2\Delta x^2} + \frac{v_{d,\mathrm{lo}}^2 + v_{d,\mathrm{up}}^2}{2\Delta y^2} \right) + \frac{g h_{d,j}}{\Delta x \Delta y}, \\ \bar{e}_k^u &= \frac{h_{d,\mathrm{le}} + h_{d,\mathrm{ri}}}{2\Delta x \Delta y} \frac{u_{d,k}}{\Delta x}, \\ \bar{e}_l^v &= \frac{h_{d,\mathrm{lo}} + h_{d,\mathrm{up}}}{2\Delta x \Delta y} \frac{v_{d,l}}{\Delta y}. \end{aligned} \tag{B.7}$$

where *le*, *ri*, *lo*, *up* refer to the left, right, lower and upper parts of the considered 2×2 control volume. We show the consistency errors for three representative cases of the sample grid shown in Fig. 3.7 on $\Omega = (0, 2\Delta x) \times (0, \frac{3\Delta y}{2}) \subset \mathbb{R}^2$. The step sizes in both directions are related via a constant $c > 0$: $\Delta y = c \Delta x$. We omit the arguments of the functions where clear from the context.

No ghost value. The numerical approximation of the discharge $e_2^u = hu$ between the faces f_1 and f_2 on the primal grid does not depend on ghost values. In order to obtain $\bar{e}_2^u\big|_*$, we substitute

$$h_{d,\mathrm{le}} + h_{d,\mathrm{ri}} = \int_{\frac{\Delta y}{2}}^{\frac{3\Delta y}{2}} \int_0^{2\Delta x} h(x, y)\, dx dy \tag{B.8}$$

and

$$u_{d,2} = \int_{\frac{\Delta x}{2}}^{\frac{3\Delta x}{2}} u(x, 0)\, dx \tag{B.9}$$

in the expression for $\bar{e}_2^u$. To determine the order of the error ϵ_2^u as defined in (3.45) for $\Delta x \to 0$, we represent

$$\lim_{\Delta x \to 0} \left(\frac{\partial e_2^u}{\partial x}\Big|_{(\Delta x, \Delta y)} - \frac{1}{\Delta x} \bar{e}_2^u\big|_* \right) \tag{B.10}$$

using Taylor series expansion of the contained terms around $(\Delta x, \Delta y)$. This local error turns out to have order $\mathcal{O}(\Delta x^2)$.

Consistent ghost values. We now consider the numerical approximation $\bar{e}_1^h$ of the hydrodynamic pressure $e_1^h = \frac{1}{2}(u^2 + v^2) + gh$ in the center of primal face f_1. If we assume zero external inflows to f_1, as indicated in Fig. 3.7, i. e.

$$\begin{aligned} e^u(0,y) = 0 \quad &\Leftrightarrow \quad u(0,y) = 0 \qquad \text{for} \qquad y \in (\frac{\Delta y}{2}, \frac{3\Delta y}{2}) \\ e^v(x, \frac{3\Delta y}{2}) = 0 \quad &\Leftrightarrow \quad v(x, \frac{3\Delta y}{2}) = 0 \qquad \text{for} \qquad x \in (0, \Delta x), \end{aligned} \tag{B.11}$$

a consistent choice of ghost velocities is $\bar{u}_{III} = 0$ and $\bar{v}_I = 0$. Accordingly, we set

$$u_{d,\text{le}} = \int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \bar{u}_{III}\, dx = 0 \qquad \text{and} \qquad v_{d,\text{up}} = \int_{\Delta y}^{2\Delta y} \bar{v}_I\, dy = 0. \tag{B.12}$$

With

$$u_{d,\text{ri}} = \int_{\frac{\Delta x}{2}}^{\frac{3\Delta x}{2}} u(x,0)\, dx, \qquad v_{d,\text{lo}} = \int_0^{\Delta y} v(0,y)\, dy \tag{B.13}$$

and

$$h_{d,1} = \int_{\frac{\Delta y}{2}}^{\frac{3\Delta y}{2}} \int_0^{\Delta x} h(x,y)\, dxdy, \tag{B.14}$$

we determine $\bar{e}_1^h\big|_*$. The Taylor series expansion of

$$\lim_{\Delta x \to 0} \left(\frac{\partial e_1^h}{\partial x}\Big|_{(\frac{\Delta x}{2}, \Delta y)} - \frac{1}{\Delta x} \bar{e}_1^h\Big|_* \right) \tag{B.15}$$

yields, besides terms of order $\mathcal{O}(\Delta x)$ and higher, constant terms, which contain $u(\frac{\Delta x}{2}, \Delta y)$ and $v(\frac{\Delta x}{2}, \Delta y)$ as factors. These factors can be developed around the points $(0, \Delta y)$ and $(\frac{\Delta x}{2}, \frac{3\Delta y}{2})$, respectively:

$$\begin{aligned} u(\frac{\Delta x}{2}, \Delta y) &= u(0, \Delta y) + \frac{\partial u}{\partial x}\bigg|_{(0,\Delta y)} \cdot \frac{\Delta x}{2} + \mathcal{O}(\Delta x^2), \\ v(\frac{\Delta x}{2}, \Delta y) &= v(\frac{\Delta x}{2}, \frac{3\Delta y}{2}) - \frac{\partial u}{\partial y}\bigg|_{(\frac{\Delta x}{2}, \frac{3\Delta y}{2})} \cdot \frac{\Delta y}{2} + \mathcal{O}(\Delta y^2). \end{aligned} \tag{B.16}$$

According to the boundary conditions (B.11), the first terms are zero, i. e. $u(0, \Delta y) = 0$ and $v(\frac{\Delta x}{2}, \frac{3\Delta y}{2}) = 0$. In this case, the error of (B.15) is of order $\mathcal{O}(\Delta x)$. The same result as presented for the error $\epsilon_1^{h,x}$ is obtained for $\epsilon_1^{h,y}$.

Inconsistent ghost values. If, other than in the previous paragraph, the choice of the ghost values is not in accordance with the boundary conditions (B.11), i. e. $\bar{u}_{III} \neq 0$ and/or $\bar{v}_I \neq 0$, the constant terms do not vanish from the error (B.15). This means that the local error for $\Delta x \to 0$ is not bounded by a function in Δx of polynomial degree greater or equal than one. The approximation of the constitutive equations on this part of the boundary is then inconsistent.

B.2 Transfer Function of the 1D Heat Equation

Consider the heat equation on $\Omega = (0,1) \subset \mathbb{R}$

$$\frac{\partial}{\partial t} x(z,t) = \frac{\partial^2}{\partial z^2} x(z,t) \tag{B.17}$$

under the Neumann-Dirichlet boundary conditions

$$\frac{\partial}{\partial z} x(0,t) = 0, \qquad x(1,t) = u(t). \tag{B.18}$$

Assuming $x(z,0) = 0$, the Laplace transform of (B.17) yields

$$s\hat{x}(z,s) = \frac{\partial^2}{\partial^2 z} \hat{x}(z,s), \tag{B.19}$$

which has the particular solution

$$\hat{x}(z,s) = c_1 e^{\sqrt{s}z} + c_2 e^{-\sqrt{s}z}, \tag{B.20}$$

with constants $c_1, c_2 \in \mathbb{R}$. Differentiation and comparison with the Neumann boundary condition yields $c_1 = c_2$. The particular solution satisfies the Dirichlet boundary condition for $c_1 = \frac{1}{e^{\sqrt{s}} + e^{-\sqrt{s}}} \hat{u}(s)$, from which we obtain

$$\hat{x}(z,s) = \frac{\cosh(\sqrt{s}z)}{\cosh(\sqrt{s})} \hat{u}(s). \tag{B.21}$$

With the output $y(t) = x(0,t)$, the transfer function is finally given by

$$\hat{y}(s) = \frac{1}{\cosh(\sqrt{s})} \hat{u}(s). \tag{B.22}$$

B.3 1D Shallow Water Equations

This sections contains the necessary considerations to design a trajectory between two stationary regimes of an open channel flow. First, the steady state solutions of the frictionless shallow water equations are determined in terms of the desired downstream water levels before and after the transient. In a second step, we verify that the desired transient corresponds to a transition between subcritical flow regimes. We consider a rectangular channel with a rising constant slope, i. e. $S_0 = \frac{b(0) - b(L)}{L} < 0$. The parameters are taken from [93], up to the sign of the bottom slope[1].

[1] The value $S_0 = 0.001$ in [93] corresponds to a falling bottom profile $b(z)$. The study of the steady state solution below, and the comparison with [93], Fig. 2, suggests that $S_0 = -0.001$, i. e. a rising channel bed is meant.

Table B.1: Parameters of the rectangular channel used in the numerical experiments according to [93], with corrected sign of S_0.

Symbol	Value	Unit
L	1000	m
S_0	-0.001	–
C	10^{-4}	$\frac{\text{s}}{\text{m}}$
R	1	–
P	1	–
m	1	–
C_w	0.4	–
w	5	m
$h_{L,1}$	8	m
$h_{L,2}$	13	m

B.3.1 Steady State Solution

To compute steady state profiles of flow depth and velocity, friction is neglected. The shallow water equations ($h = p$, $v = q$) in steady state with a bed elevation $b(z) = S_0(L - z)$ are

$$0 = -\frac{\partial}{\partial z}(h(z)v(z)) \tag{B.23a}$$

$$0 = -\frac{\partial}{\partial z}(\frac{1}{2}v^2(z) + g(h(z) + b(z))). \tag{B.23b}$$

From the boundary condition (6.66) at the downstream overflow weir (height w), we can determine the flow velocity based on a desired flow depth (notation $h_L = h(L)$, $v_L = v(L)$):

$$v_L = C_w \frac{2}{3}\sqrt{2g}\frac{(h_L - w)^{\frac{3}{2}}}{h_L}. \tag{B.24}$$

The conservation of mass (B.23a) and momentum (B.23b) imply

$$h(z)v(z) = h_L v_L \tag{B.25a}$$

$$\frac{1}{2g}v^2(z) + h(z) + b(z) = \frac{1}{2g}v_L^2 + h_L. \tag{B.25b}$$

With

$$v(z) = \frac{h_L v_L}{h(z)} = \frac{Q_L}{h(z)}, \tag{B.26}$$

where Q_L denotes the constant discharge, the second equation can be written in terms of $h(z)$ as the only unknown.

$$\frac{1}{2g}\frac{h_L^2 v_L^2}{h^2(z)} + h(z) + b(z) = \frac{1}{2g}v_L^2 + h_L. \tag{B.27}$$

The numerically computed solution $h(z)$ as well as the resulting velocity profile $v(z)$ for the parameters given in Table B.1 are depicted in Fig. B.1.

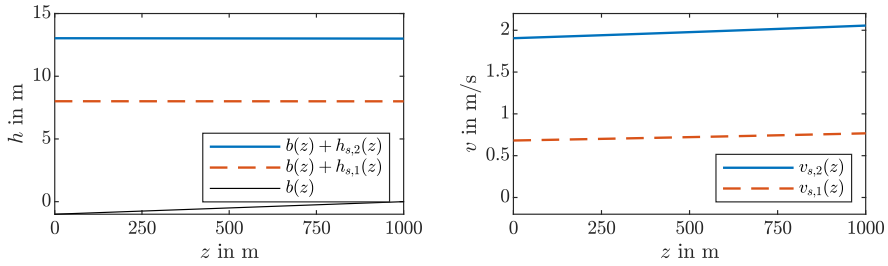

Figure B.1: Steady state water level and velocity profile before (index 1) and after (index 2) the transient, based on the frictionless case.

B.3.2 Flow Regime

To determine the flow regimes, which correspond to the steady states before and after the transient, we rewrite (B.27) in terms of the steady state discharge,

$$\underbrace{\frac{1}{2g}\frac{Q_L^2}{h^2(z)} + h(z)}_{E(h(z))} + b(z) = \underbrace{\frac{1}{2g}\frac{Q_L^2}{h_L^2} + h_L}_{E(h_L)}. \tag{B.28}$$

$E(h(z))$ and $E(h_L)$ denote *specific energies*, which represent the *total head* above the channel bottom for a given discharge Q_L, see [38], p. 31. The specific energy $E(h)$ is a function which tends to ∞ for $h \to 0$ and which approaches h for $h \to \infty$. Its minimum, given by

$$\frac{\partial E(h)}{\partial h} = 0 \qquad \Leftrightarrow \qquad -\frac{Q_L^2}{gh^3} + 1 = 0, \tag{B.29}$$

defines the *critical depth*

$$h_c = \sqrt[3]{\frac{Q_L^2}{g}}. \tag{B.30}$$

Combining (B.26) with (B.30) gives the critical velocity, which corresponds to the critical depth

$$v_c = \sqrt{gh_c}. \tag{B.31}$$

For water depth $h(z) > h_c$, the flow is referred to as *subcritical*, for $h(z) < h_c$ as *supercritical*[2]. The desired transition between the steady state dowstream depths $h_{L,1} = 8\,\text{m}$ and $h_{L,2} = 13\,\text{m}$ takes place in a subcritical flow regime, as illustrated in the right diagram of Fig. B.2.

With $E(h(z)) \geq E_{\min}$, the inequality

$$E(h_L) \geq E_{\min} + b(z) \tag{B.32}$$

[2]Hydraulic jumps occur at the transition between supercritical and subcritical flow, see [38], Section 2-8.

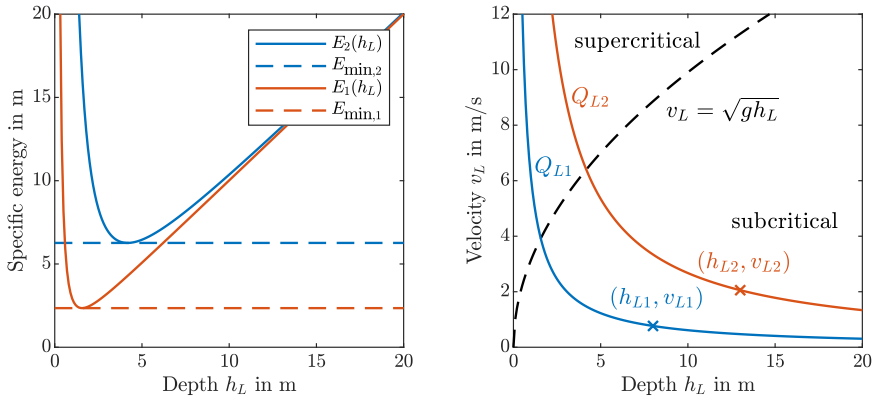

Figure B.2: Left: Specific energies and their minimum for both steady state discharges $Q_{L1} < Q_{L2}$. Right: Flow regimes and curves of constant discharge.

follows from (B.28), and is certainly true if the downstream specific energy satisfies

$$E(h_L) \geq E_{\min} + \max_z b(z). \tag{B.33}$$

The validity of this inequality can be verified for the desired downstream discharges and channel parameters.

Bibliography

[1] J. A. Acosta, R. Ortega, A. Astolfi, and A. D. Mahindrakar. Interconnection and Damping Assignment Passivity-Based Control of mechanical systems with underactuation degree one. *IEEE Transactions on Automatic Control*, 50(12):1936–1955, 2005.

[2] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The FEniCS project version 1.5. *Archive of Numerical Software*, 3(100):9–23, 2015.

[3] R. Altmann and P. Schulze. A port-Hamiltonian formulation of the Navier-Stokes equations for reactive flows. *Systems & Control Letters*, 100:51–55, 2017.

[4] S. Aoues, D. Eberard, and W. Marquis-Favre. Canonical interconnection of discrete linear port-Hamiltonian systems. In *52nd IEEE Conference on Decision and Control, Florence*, pages 3166–3171, 2013.

[5] A. Arakawa and V. R. Lamb. A potential enstrophy and energy conserving scheme for the shallow water equations. *Monthly Weather Review*, 109(1):18–36, 1981.

[6] D. Arnold, R. Falk, and R. Winther. Finite element exterior calculus: from Hodge theory to numerical stability. *Bulletin of the American Mathematical Society*, 47(2):281–354, 2010.

[7] D. N. Arnold. Spaces of finite element differential forms. In *Analysis and Numerics of Partial Differential Equations*, pages 117–140. Springer, 2013.

[8] V. I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60. Springer, 1989.

[9] A. Baaiu, F. Couenne, D. Eberard, C. Jallut, L. Lefèvre, Y. Le Gorrec, and B. Maschke. Port-based modelling of mass transport phenomena. *Mathematical and Computer Modelling of Dynamical Systems*, 15(3):233–254, 2009.

[10] A. Baaiu, F. Couenne, L. Lefevre, Y. Le Gorrec, and M. Tayakout. Structure-preserving infinite dimensional model reduction: Application to adsorption processes. *Journal of Process Control*, 19(3):394–404, 2009.

[11] G. Bastin and J.-M. Coron. *Stability and Boundary Stabilization of 1-D Hyperbolic Systems*, volume 88. Birkhäuser, 2016.

[12] C. Batlle, A. Dòria-Cerezo, G. Espinosa-Pérez, and R. Ortega. Simultaneous Interconnection and Damping Assignment Passivity-Based Control: The induction machine case study. *International Journal of Control*, 82(2):241–255, 2009.

[13] C. Beattie, V. Mehrmann, H. Xu, and H. Zwart. Linear port-Hamiltonian

descriptor systems. *Mathematics of Control, Signals, and Systems*, 30(4):17, 2018.

[14] J. Becker and T. Meurer. Feedforward tracking control for non-uniform Timoshenko beam models: combining differential flatness, modal analysis, and FEM. *ZAMM – Journal of Applied Mathematics and Mechanics*, 87(1):37–58, 2007.

[15] G. Blankenstein, R. Ortega, and A. J. van der Schaft. The matching conditions of Controlled Lagrangians and IDA-Passivity Based Control. *International Journal of Control*, 75(9):645–665, 2002.

[16] A. M. Bloch, D. E. Chang, N. E. Leonard, and J. E. Marsden. Controlled Lagrangians and the stabilization of mechanical systems II: Potential shaping. *IEEE Transactions on Automatic Control*, 46:1556–1571, 2001.

[17] A. M. Bloch, N. E. Leonard, and J. E. Marsden. Controlled Lagrangians and the stabilization of mechanical systems I: The first matching theorem. *IEEE Transactions on Automatic Control*, 45(12):2253–2270, 2000.

[18] P. B. Bochev and J. M. Hyman. Principles of mimetic discretizations of differential operators. In *Compatible spatial discretizations*, pages 89–119. Springer, 2006.

[19] T. Böhm and T. Meurer. Trajectory planning and tracking control for the temperature distribution in a deep drawing tool. *Control Engineering Practice*, 64:127–139, 2017.

[20] A. Bossavit. Differential forms and the computation of fields and forces in electromagnetism. *European Journal of Mechanics – B/Fluids*, 10(5):474–488, 1991.

[21] A. Bossavit. *Computational Electromagnetism: Variational Formulations, Complementarity, Edge Elements*. Academic Press, 1998.

[22] A. Bossavit. How weak is the "weak solution" in finite element methods? *IEEE Transactions on Magnetics*, 34(5):2429–2432, 1998.

[23] A. Bossavit. Generating Whitney forms of polynomial degree one and higher. *IEEE Transactions on Magnetics*, 38(2):341–344, 2002.

[24] A. Bossavit and L. Kettunen. Yee-like schemes on a tetrahedral mesh, with diagonal lumping. *International Journal of Numerical Modelling Electronic Networks Devices and Fields*, 12:129–142, 1999.

[25] P. C. Breedveld. Multibond graph elements in physical systems theory. *Journal of the Franklin Institute*, 319(1-2):1–36, 1985.

[26] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011.

[27] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*, volume 15. Springer, 1991.

[28] T. J. Bridges and S. Reich. Multi-symplectic integrators: numerical schemes for Hamiltonian PDEs that conserve symplecticity. *Physics Letters A*, 284(4):184–193, 2001.

[29] A. Brugnoli, D. Alazard, V. Pommier-Budinger, and D. Matignon. Port-Hamiltonian formulation and symplectic discretization of plate models. Part I: Mindlin model for thick plates. *Applied Mathematical Modeling*, 75:940–960, 2019.

[30] A. Brugnoli, D. Alazard, V. Pommier-Budinger, and D. Matignon. Port-Hamiltonian formulation and symplectic discretization of plate models. Part II: Kirchhoff model for thin plates. *Applied Mathematical Modeling*, 75:961–981, 2019.

[31] C. I. Byrnes, A. Isidori, and J. C. Willems. Passivity, feedback equivalence, and the global stabilization of minimum phase nonlinear systems. *IEEE Transactions on Automatic Control*, 36:1228–1240, 1991.

[32] R. Camassa, G. Falqui, G. Ortenzi, and M. Pedroni. On variational formulations and conservation laws for incompressible 2D Euler fluids. *Journal of Physics: Conference Series*, 482(1):012006, 2014.

[33] F. L. Cardoso Ribeiro. *Port-Hamiltonian modeling and control of a fluid-structure system.* PhD thesis, Université de Toulouse, 2016.

[34] F. L. Cardoso-Ribeiro, D. Matignon, and L. Lefèvre. A structure-preserving partitioned finite element method for the 2D wave equation. *IFAC-PapersOnLine*, 51(3):119–124, 2018.

[35] F. L. Cardoso-Ribeiro, D. Matignon, and V. Pommier-Budinger. A port-Hamiltonian model of liquid sloshing in moving containers and application to a fluid-structure system. *Journal of Fluids and Structures*, 69:402–427, 2017.

[36] E. Celledoni and E. H. Høiseth. Energy-preserving and passivity-consistent numerical discretization of port-Hamiltonian systems. *arXiv preprint arXiv:1706.08621*, 2017.

[37] J. Cervera, A. J. van der Schaft, and A. Baños. Interconnection of port-Hamiltonian systems and composition of Dirac structures. *Automatica*, 43(2):212–225, 2007.

[38] M. H. Chaudhry. *Open-Channel Flow.* Prentice Hall, 1993.

[39] S. H. Christiansen. Upwinding in finite element systems of differential forms. In *Foundations of Computational Mathematics, Budapest 2011*, volume 403 of *London Math. Soc. Lecture Note Ser.* Cambridge Univ. Press, 2013.

[40] D. Cohen and E. Hairer. Linear energy-preserving integrators for Poisson systems. *BIT Numerical Mathematics*, 51(1):91–101, 2011.

[41] C. Cotter and J. Shipton. Mixed finite elements for numerical weather prediction. *Journal of Computational Physics*, 231(21):7076–7091, 2012.

[42] C. J. Cotter and J. Thuburn. A finite element exterior calculus framework for the rotating shallow-water equations. *Journal of Computational Physics*, 257:1506–1526, 2014.

[43] T. J. Courant. Dirac manifolds. *Transactions of the American Mathematical Society*, 319(2):631–661, 1990.

[44] R. F. Curtain and H. Zwart. *An Introduction to Infinite-dimensional Linear Systems Theory.* Springer, 1995.

[45] C. M. Dafermos. *Hyperbolic Conservation Laws in Continuum Physics.* Springer, 2010.

[46] S. Delgado and P. Kotyczka. Energy shaping for position and speed control of a wheeled inverted pendulum. *Automatica*, 74:222–229, 2016.

[47] M. Desbrun, A. N. Hirani, M. Leok, and J. E. Marsden. Discrete exterior calculus. *arXiv preprint math/0508341*, 2005.

[48] M. Desbrun, E. Kanso, and Y. Tong. Discrete differential forms for computational modeling. In *Discrete Differential Geometry*, pages 287–324. Springer, 2008.

[49] S. Devasia, D. Chen, and B. Paden. Nonlinear inversion-based output tracking. *IEEE Transactions on Automatic Control*, 41(7):930–942, 1996.

[50] V. Duindam, A. Macchelli, S. Stramigioli, and H. Bruyninckx. *Modeling and Control of Complex Physical Systems: The Port-Hamiltonian Approach*. Springer, 2009.

[51] D. Eberard, B. M. Maschke, and A. J. van der Schaft. An extension of Hamiltonian systems to the thermodynamic phase space: Towards a geometry of nonreversible processes. *Reports on Mathematical Physics*, 60(2):175–198, 2007.

[52] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. *Handbook of Numerical Analysis*, 7:713–1018, 2000.

[53] R. Eymard and R. Herbin. A staggered finite volume scheme on general meshes for the Navier-Stokes equations in two space dimensions. *International Journal on Finite Volumes*, 2(1):1–18, 2005.

[54] A. Falaize and T. Hélie. Passive guaranteed simulation of analog audio circuits: a port-Hamiltonian approach. *Applied Sciences*, 6(10):273, 2016.

[55] O. Farle, R.-B. Baltes, and R. Dyczij-Edlinger. Strukturerhaltende Diskretisierung verteilt-parametrischer Port-Hamiltonscher Systeme mittels finiter Elemente. *at – Automatisierungstechnik*, 62(7):500–511, 2014.

[56] O. Farle, D. Klis, M. Jochum, O. Floch, and R. Dyczij-Edlinger. A port-Hamiltonian finite-element formulation for the Maxwell equations. In *International Conference on Electromagnetics in Advanced Applications*, pages 324–327, 2013.

[57] H. O. Fattorini. Boundary control systems. *SIAM Journal on Control*, 6(3):349–385, 1968.

[58] S. Fiaz, D. Zonetti, R. Ortega, J. M. A. Scherpen, and A. J. van der Schaft. A port-Hamiltonian approach to power network modeling and analysis. *European Journal of Control*, 19(6):477–485, 2013.

[59] H. Flanders. *Differential Forms with Applications to the Physical Sciences*. Academic Press, New York, 1963.

[60] M. Fliess, J. Lévine, P. Martin, and P. Rouchon. Flatness and defect of nonlinear systems: Introductory theory and examples. *International Journal of Control*, 61(6):1327–1361, 1995.

[61] M. Fliess, P. Martin, N. Petit, and P. Rouchon. Active signal restoration for the telegraph equation. In *38th IEEE Conference on Decision and Control, Phoenix*, volume 2, pages 1107–1111, 1999.

[62] M. Fliess, H. Mounier, P. Rouchon, and J. Rudolph. Systèmes linéaires sur les opérateurs de Mikusinski et commande d'une poutre flexible. In *ESAIM: Proceedings*, volume 2, pages 183–193, 1997.

[63] M. Fliess, H. Mounier, P. Rouchon, and J. Rudolph. A distributed parameter approach to the control of a tubular reactor: A multivariable case. In *37th IEEE Conference on Decision and Control, Tampa*, volume 1, pages 439–442, 1998.

[64] B. Fornberg. High-order finite differences and the pseudospectral method on staggered grids. *SIAM Journal on Numerical Analysis*, 27(4):904–918, 1990.

[65] T. Frankel. *The Geometry of Physics: An Introduction*. Cambridge University Press, 2011.

[66] P. J. Gawthrop and G. P. Bevan. Bond-graph modeling – a tutorial introduction for control engineers. *IEEE Control Systems Magazine*, pages 24–45, 2007.

[67] C. Geuzaine. GetDP: a general finite-element solver for the de Rham complex. *Proceedings in Applied Mathematics and Mechanics*, 7:1010603–1010604, 2008.

[68] M. Gevrey. Sur la nature analytique des solutions des équations aux dérivées partielles. premier mémoire. In *Annales Scientifiques de l'Ecole Normale Supérieure*, volume 35, pages 129–190, 1918.

[69] M. Giftthaler, T. Wolf, H. K. F. Panzer, and B. Lohmann. Parametric model order reduction of port-Hamiltonian systems by matrix interpolation. *at – Automatisierungstechnik*, 62(9):619–628, 2014.

[70] G. Golo. *Interconnection structures in port-based modelling: tools for analysis and simulation*. PhD thesis, Universiteit Twente, 2002.

[71] G. Golo, V. Talasila, A. van der Schaft, and B. Maschke. Hamiltonian discretization of boundary control systems. *Automatica*, 40(5):757–771, 2004.

[72] L. Gören Sümer and Y. Yalçin. Gradient based discrete-time modeling and control of Hamiltonian systems. *IFAC Proceedings Volumes*, 41(2):212–217, 2008.

[73] L. Gören Sümer and Y. Yalçin. A direct discrete-time IDA-PBC design method for a class of underactuated Hamiltonian systems. In *18th IFAC World Congress, Milano*, pages 13456–13461, 2011.

[74] S. Gugercin, R. V. Polyuga, C. Beattie, and A. J. van der Schaft. Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. *Automatica*, 48(9):1963–1974, 2012.

[75] G. Haine et al. INFIDHEM project website. https://websites.isae-supaero.fr/infidhem/the-project/.

[76] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31. Springer, 2006.

[77] B. Hamroun. *Approche hamiltonienne à ports pour la modélisation, la réduction et la commande des systèmes non linéaires à paramètres distribués – Application aux écoulements à surface libre*. PhD thesis, Grenoble INP, 2009.

[78] B. Hamroun, A. Dimofte, L. Lefèvre, and E. Mendes. Control by Interconnection and energy-shaping methods of port Hamiltonian models. Application to the shallow water equations. *European Journal of Control*, 16(5):545–563, 2010.

[79] F. Hecht. New development in FreeFem++. *Journal of Numerical Mathematics*, 20(3-4):251–265, 2012.

[80] J. Henikl, W. Kemmetmüller, T. Meurer, and A. Kugi. Infinite-dimensional decentralized damping control of large-scale manipulators with hydraulic actuation. *Automatica*, 63:101–115, 2016.

[81] J. S. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer, 2007.

[82] R. Hiemstra, D. Toshniwal, R. Huijsmans, and M. I. Gerritsma. High order geometric methods with exact conservation properties. *Journal of Computational Physics*, 257:1444–1471, 2014.

[83] R. Hiptmair. Finite elements in computational electromagnetism. *Acta Numerica*, 11:237–339, 2002.

[84] A. N. Hirani. *Discrete exterior calculus*. PhD thesis, California Institute of Technology, 2003.

[85] H. Hoang, F. Couenne, C. Jallut, and Y. Le Gorrec. The port Hamiltonian approach to modeling and control of continuous stirred tank reactors. *Journal of Process Control*, 21(10):1449–1458, 2011.

[86] D. D. Holm. *Geometric Mechanics, Part 1: Dynamics and Symmetry*. Imperial College Press, 2nd edition, 2011.

[87] A. Iserles. Generalized leapfrog methods. *IMA Journal of Numerical Analysis*, 6(4):381–392, 1986.

[88] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University Press, 2009.

[89] B. Jacob and H. Zwart. *Linear Port-Hamiltonian Systems on Infinite-dimensional Spaces*, volume 223. Springer, 2012.

[90] K. Jänich. *Vector Analysis*. Springer, 2001.

[91] A. Kaldmäe and Ü. Kotta. On flatness of discrete-time nonlinear systems. *IFAC Proceedings Volumes*, 46(23):588–593, 2013.

[92] T. Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer, 1995.

[93] T. Knüppel, F. Woittennek, and J. Rudolph. Flatness-based trajectory planning for the shallow water equations. In *49th IEEE Conference on Decision and Control, Atlanta*, pages 2960–2965, 2010.

[94] P. Kotyczka. Local linear dynamics assignment in IDA-PBC. *Automatica*, 49(4):1037–1044, 2013.

[95] P. Kotyczka. Finite volume structure-preserving discretization of 1D distributed-parameter port-Hamiltonian systems. *IFAC-PapersOnLine*, 49(8):298–303, 2016.

[96] P. Kotyczka. Structured discretization of the heat equation: Numerical properties and preservation of flatness. In *23rd International Symposium on Mathematical Theory of Networks and Systems, Hong Kong*, pages 600–607, 2018.

[97] P. Kotyczka. Zur Erhaltung von Struktur und Flachheit bei der torbasierten Ortsdiskretisierung. *at – Automatisierungstechnik*, 66(7):511–535, 2018.

[98] P. Kotyczka. Discrete-time flatness-based feedforward control for the 1D shallow water equations. In *Joint 8th IFAC Symposium on Mechatronic Systems and 11th IFAC Symposium on Nonlinear Control Systems, Vienna*, pages 44–49, 2019.

[99] P. Kotyczka and A. Blancato. Feedforward control of a channel flow based on a discretized port-Hamiltonian model. *IFAC-PapersOnLine*, 48(13):194–199, 2015.

[100] P. Kotyczka, H. Joos, Y. Wu, and Y. Le Gorrec. Finite-dimensional observers for port-Hamiltonian systems of conservation laws. In *58th IEEE Conference on Decision and Control, Nice*, 2019.

[101] P. Kotyczka and L. Lefèvre. Discrete-time port-Hamiltonian systems based on Gauss-Legendre collocation. *IFAC-PapersOnLine*, 51(3):125–130, 2018.

[102] P. Kotyczka and L. Lefèvre. Discrete-time port-Hamiltonian systems: A definition based on symplectic integration. *Systems & Control Letters*, 133:104530, 2019.

[103] P. Kotyczka and B. Maschke. Discrete port-Hamiltonian formulation and numerical approximation for systems of two conservation laws. *at – Automatisierungstechnik*, 65(5):308–322, 2017.

[104] P. Kotyczka, B. Maschke, and L. Lefèvre. Weak form of Stokes-Dirac structures and geometric discretization of port-Hamiltonian systems. *Journal of Computational Physics*, 361:442–476, 2018.

[105] J. Kreeft and M. Gerritsma. Mixed mimetic spectral element method for Stokes flow: A pointwise divergence-free solution. *Journal of Computational Physics*, 240:284–309, 2013.

[106] M. Kurula and H. Zwart. Linear wave systems on n-d spatial domains. *International Journal of Control*, 88(5):1063–1077, 2015.

[107] M. Kurula, H. Zwart, A. J. van der Schaft, and J. Behrndt. Dirac structures and their composition on Hilbert spaces. *Journal of Mathematical Analysis and Applications*, 372(2):402–422, 2010.

[108] D. S. Laila and A. Astolfi. Discrete-time IDA-PBC design for separable Hamiltonian systems. In *16th IFAC World Congress, Prague*, pages 838–843, 2005.

[109] B. Laroche, P. Martin, P. Rouchon, et al. Motion planning for the heat equation. *International Journal of Robust and Nonlinear Control*, 10(8):629–643, 2000.

[110] M. G. Larson and F. Bengzon. *The Finite Element Method: Theory, Implementation, and Applications*, volume 10. Springer, 2013.

[111] Y. Le Gorrec, H. Peng, L. Lefèvre, B. Hamroun, and F. Couenne. Systèmes hamiltoniens à ports de dimension infinie: réduction et propriétés spectrales. *Journal Européen des Systèmes Automatisés*, 45(7-10):645–664, 2011.

[112] Y. Le Gorrec, H. Zwart, and B. Maschke. Dirac structures and boundary control systems associated with skew-symmetric differential operators. *SIAM Journal on Control and Optimization*, 44(5):1864–1892, 2005.

[113] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*, volume 14. Cambridge University Press, 2004.

[114] R. J. LeVeque. *Numerical Methods for Conservation Laws*, volume 132. Springer, 1992.

[115] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*, volume 31. Cambridge University Press, 2002.

[116] J. Lévine. *Analysis and Control of Nonlinear Systems*. Springer, 2009.

[117] J. Lévine. On necessary and sufficient conditions for differential flatness. *Applicable Algebra in Engineering, Communication and Computing*, 22(1):47–90, 2011.

[118] A. Lew, J. E. Marsden, M. Ortiz, and M. West. An overview of variational integrators. In *Finite Element Methods: 1970's and Beyond. Theory and engineering applications of computational methods*, pages 1–18. International Center for Numerical Methods in Engineering (CIMNE), Barcelona, 2004.

[119] Z.-H. Luo, B.-Z. Guo, and Ö. Morgül. *Stability and Stabilization of Infinite Dimensional Systems with Applications*. Springer, 1999.

[120] A. F. Lynch and J. Rudolph. Flachheitsbasierte Randsteuerung parabolischer Systeme mit verteilten Parametern. *at – Automatisierungstechnik*, 48(10):478–486, 2000.

[121] A. F. Lynch and D. Wang. Flatness-based control of a flexible beam in a gravitational field. In *American Control Conference, Boston*, pages 5449–5455, 2004.

[122] A. Macchelli. Energy shaping of distributed parameter port-Hamiltonian systems based on finite element approximation. *Systems & Control Letters*, 60:579–589, 2011.

[123] A. Macchelli. Dirac structures on Hilbert spaces and boundary control of distributed port-Hamiltonian systems. *Systems & Control Letters*, 68:43–50, 2014.

[124] A. Macchelli. Passivity-based control of implicit port-Hamiltonian systems. *SIAM Journal on Control and Optimization*, 52(4):2422–2448, 2014.

[125] A. Macchelli, Y. Le Gorrec, H. Ramírez, and H. Zwart. On the synthesis of boundary control laws for distributed port-Hamiltonian systems. *IEEE Transactions on Automatic Control*, 62(4):1700–1713, 2017.

[126] A. Macchelli and C. Melchiorri. Modeling and control of the Timoshenko beam. The distributed port Hamiltonian approach. *SIAM Journal on Control and Optimization*, 43(2):743–767, 2004.

[127] A. Macchelli and C. Melchiorri. Control by interconnection of mixed port Hamiltonian systems. *IEEE Transactions on Automatic Control*, 50(11):1839–1844, 2005.

[128] A. Macchelli, C. Melchiorri, and L. Bassi. Port-based modelling and control of the Mindlin plate. In *44th IEEE Conference on Decision and Control and European Control Conference, Sevilla*, pages 5989–5994, 2005.

[129] A. Macchelli, C. Melchiorri, and S. Stramigioli. Port-based modeling of a flexible link. *IEEE Transactions on Robotics*, 23(4):650–660, 2007.

[130] B. Maschke, R. Ortega, and A. J. van der Schaft. Energy-based Lyapunov functions for forced Hamiltonian systems with dissipation. *IEEE Transactions on Automatic Control*, 45(8):1498–1502, 2000.

[131] B. Maschke, A. J. van der Schaft, and P. C. Breedveld. An intrinsic Hamiltonian formulation of network dynamics: Non-standard Poisson structures and gyrators. *Journal of the Franklin Institute*, 329(5):923–966, 1992.

[132] B. M. Maschke and A. J. van der Schaft. Port-controlled Hamiltonian systems: modelling origins and systemtheoretic properties. *IFAC Proceedings Volumes*, 25(13):359–365, 1992.

[133] T. Meurer. Flatness-based trajectory planning for diffusion-reaction systems in a parallelepipedon – a spectral approach. *Automatica*, 47(5):935–949, 2011.

[134] T. Meurer. *Control of Higher–Dimensional PDEs: Flatness and Backstepping Designs*. Springer, 2012.

[135] T. Meurer and M. Zeitz. Flachheitsbasierte Steuerung und Regelung eines Wärmeleitungssystems. *at – Automatisierungstechnik*, 52(9):411–420, 2004.

[136] M. Miletic, D. Stürzer, A. Arnold, and A. Kugi. Stability of an Euler-Bernoulli

beam with a nonlinear dynamic feedback system. *IEEE Transactions on Automatic Control*, 5:6, 2016.

[137] P. J. Morrison. Hamiltonian description of the ideal fluid. *Reviews of Modern Physics*, 70(2):467–521, 1998.

[138] R. Moulla, L. Lefèvre, and B. Maschke. Pseudo-spectral methods for the spatial symplectic reduction of open systems of conservation laws. *Journal of Computational Physics*, 231(4):1272–1292, 2012.

[139] H. Mounier and J. Rudolph. Flatness-based control of nonlinear delay systems: A chemical reactor example. *International Journal of Control*, 71(5):871–890, 1998.

[140] G. Nishida, B. Maschke, and R. Ikeura. Boundary integrability of multiple Stokes-Dirac structures. *SIAM Journal on Control and Optimization*, 53(2):800–815, 2015.

[141] G. Nishida, K. Takagi, B. Maschke, and T. Osada. Multi-scale distributed parameter modeling of ionic polymer-metal composite soft actuator. *Control Engineering Practice*, 19(4):321–334, 2011.

[142] G. Nishida and M. Yamakita. Distributed port Hamiltonian formulation of flexible beams under large deformations. In *IEEE Conference on Control Applications, Toronto*, pages 589–594, 2005.

[143] F. Ollivier and A. Sedoglavic. A generalization of flatness to nonlinear systems of partial differential equations. Application to the command of a flexible rod. *IFAC Proceedings Volumes*, 34(6):219–223, 2001.

[144] P. J. Olver. *Introduction to Partial Differential Equations*. Springer, 2014.

[145] R. Ortega and E. García-Canseco. Interconnection and Damping Assignment Passivity-Based Control: A survey. *European Journal of Control*, 10(5):432–450, 2004.

[146] R. Ortega, A. Loría, P. J. Nicklasson, and H. Sira-Ramírez. *Passivity-based Control of Euler-Lagrange Systems*. Springer, 1998.

[147] R. Ortega, M. W. Spong, F. Gómez-Estern, and G. Blankenstein. Stabilization of a class of underactuated mechanical systems via Interconnection and Damping Assignment. *IEEE Transactions on Automatic Control*, 47(8):1218–1233, 2002.

[148] R. Ortega, A. van der Schaft, B. Maschke, and G. Escobar. Interconnection and Damping Assignment Passivity-Based control of port-controlled Hamiltonian systems. *Automatica*, 38(4):585–596, 2002.

[149] R. Ortega, A. J. van der Schaft, F. Castaños, and A. Astolfi. Control by Interconnection and standard passivity-based control of port-Hamiltonian systems. *IEEE Transactions on Automatic Control*, 53(11):2527–2542, 2008.

[150] R. Ortega, A. J. van der Schaft, I. Mareels, and B. Maschke. Putting energy back in control. *IEEE Control Systems Magazine*, 21:18–33, 2001.

[151] R. Pasumarthy, V. R. Ambati, and A. J. van der Schaft. Port-Hamiltonian formulation of shallow water equations with Coriolis force and topography. In *18th International Symposium on Mathematical Theory of Networks and Systems, Blacksburg*, 2008.

[152] R. Pasumarthy and A. van der Schaft. Hamiltonian formulation of two dimen-

sional shallow water flows with boundary energy flow. In *20th Int. Symposium on Mathematical Theory of Networks and Systems, Melbourne*, 2012.

[153] S. Patankar. *Numerical Heat Transfer and Fluid Flow*. CRC Press, 1980.

[154] H. M. Paynter. *Analysis and Design of Engineering Systems*. MIT Press, 1961.

[155] M. Polner and J. van der Vegt. A Hamiltonian vorticity–dilatation formulation of the compressible Euler equations. *Nonlinear Analysis: Theory, Methods & Applications*, 109:113–135, 2014.

[156] R. V. Polyuga and A. J. van der Schaft. Structure preserving model reduction of port-Hamiltonian systems by moment matching at infinity. *Automatica*, 46(4):665–672, 2010.

[157] S. Prajna, A. J. van der Schaft, and G. Meinsma. An LMI approach to stabilization of linear port-controlled Hamiltonian systems. *Systems & Control Letters*, 45:371–385, 2002.

[158] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer, 1994.

[159] T. S. Rabbani, F. Di Meglio, X. Litrico, and A. M. Bayen. Feed-forward control of open channel flow using differential flatness. *IEEE Transactions on Control Systems Technology*, 18(1):213–221, 2010.

[160] H. Ramírez, Y. Le Gorrec, A. Macchelli, and H. Zwart. Exponential stabilization of boundary controlled port-Hamiltonian systems with dynamic feedback. *IEEE Transactions on Automatic Control*, 59(10):2849–2855, 2014.

[161] H. Ramírez, Y. Le Gorrec, B. Maschke, and F. Couenne. On the passivity based control of irreversible processes: A port-Hamiltonian approach. *Automatica*, 64:105–111, 2016.

[162] H. Ramírez, B. Maschke, and D. Sbarbaro. Irreversible port-Hamiltonian systems: A general formulation of irreversible processes with application to the CSTR. *Chemical Engineering Science*, 89:223–234, 2013.

[163] F. Rapetti and A. Bossavit. Whitney forms of higher degree. *SIAM Journal on Numerical Analysis*, 47(3):2369–2386, 2009.

[164] S. Reich. Finite volume methods for multi-symplectic PDEs. *BIT Numerical Mathematics*, 40(3):559–582, 2000.

[165] T. Ringler, J. Thuburn, J. B. Klemp, and W. C. Skamarock. A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids. *Journal of Computational Physics*, 229(9):3065–3090, 2010.

[166] J. Rudolph. *Flatness Based Control of Distributed Parameter Systems*. Shaker, 2003.

[167] J. Rudolph, J. Winkler, and F. Woittennek. *Flatness Based Control of Distributed Parameter Systems: Examples and Computer Exercises from Various Technological Domains*. Shaker, 2003.

[168] T. Scheuermann, P. Kotyczka, M.-L. Zanota, I. Pitault, and B. Maschke. Numerical approximation of heat transfer on heterogeneous media. In *90th GAMM Annual Meeting, Vienna, Austria, Proceedings in Applied Mathematics and Mechanics (PAMM)*, 2019.

[169] M. Schöberl and K. Schlacher. Variational principles for different representations of Lagrangian and Hamiltonian systems. In *Dynamics and Control of Advanced Structures and Machines*, pages 65–73. Springer, 2017.

[170] M. Schöberl and A. Siuka. Analysis and comparison of port-Hamiltonian formulations for field theories-demonstrated by means of the Mindlin plate. In *European Control Conference, Zurich*, pages 548–553, 2013.

[171] M. Schöberl and A. Siuka. On Casimir functionals for infinite-dimensional port-Hamiltonian control systems. *IEEE Transactions on Automatic Control*, 58(7):1823–1828, 2013.

[172] M. Schöberl and A. Siuka. Jet bundle formulation of infinite-dimensional port-Hamiltonian systems using differential operators. *Automatica*, 50(2):607–613, 2014.

[173] A. Serhani, D. Matignon, and G. Haine. Structure-preserving finite volume method for 2D linear and non-linear port-Hamiltonian systems. *IFAC-PapersOnLine*, 51(3):131–136, 2018.

[174] M. Seslija, J. M. A. Scherpen, and A. J. van der Schaft. Explicit simplicial discretization of distributed-parameter port-Hamiltonian systems. *Automatica*, 50(2):369–377, 2014.

[175] M. Seslija, A. van der Schaft, and J. M. Scherpen. Discrete exterior geometry approach to structure-preserving discretization of distributed-parameter port-Hamiltonian systems. *Journal of Geometry and Physics*, 62(6):1509–1531, 2012.

[176] H. Sira-Ramírez and S. K. Agrawal. *Differentially Flat Systems*. Marcel Dekker, Inc., 2004.

[177] A. Siuka, M. Schöberl, and K. Schlacher. Port-Hamiltonian modelling and energy-based control of the Timoshenko beam. *Acta Mechanica*, 222(1):69–89, 2011.

[178] R. Specogna. Diagonal discrete Hodge operators for simplicial meshes using the signed dual complex. *IEEE Transactions on Magnetics*, 51(3):1–4, 2015.

[179] S. Stramigioli, C. Secchi, A. J. van der Schaft, and C. Fantuzzi. Sampled data systems passivity and discrete port-Hamiltonian systems. *IEEE Transactions on Robotics*, 21(4):574–587, 2005.

[180] G. Sun. Construction of high order symplectic Runge-Kutta methods. *Journal of Computational Mathematics*, 11(3):250–260, 1993.

[181] R. Szymkiewicz. Solution of the inverse problem for the Saint Venant equations. *Journal of Hydrology*, 147(1–4):105–120, 1993.

[182] V. Talasila, J. Clemente-Gallardo, and A. J. van der Schaft. Discrete port-Hamiltonian systems. *Systems & Control Letters*, 55(6):478–486, 2006.

[183] B. D. H. Tellegen. A general network theorem, with applications. *Philips Research Reports*, 7:256–269, 1952.

[184] F. Tiefensee, S. Monaco, and D. Normand-Cyrot. IDA-PBC under sampling for Port-Controlled Hamiltonian systems. In *American Control Conference, Baltimore*, pages 1811–1816, 2010.

[185] E. Tonti. A direct discrete formulation of field laws: The cell method. *CMES – Computer Modeling in Engineering and Sciences*, 2(2):237–258, 2001.

[186] V. Trenchant, Y. Fares, H. Ramírez, and Y. Le Gorrec. A port-Hamiltonian formulation of a 2D boundary controlled acoustic system. *IFAC-PapersOnLine*, 48(13):235–240, 2015.

[187] V. Trenchant, H. Ramírez, Y. Le Gorrec, and P. Kotyczka. Structure preserv-

ing spatial discretization of 2D hyperbolic systems using staggered grids finite difference. In *American Control Conference, Seattle*, pages 2491–2496, 2017.

[188] V. Trenchant, H. Ramírez, Y. Le Gorrec, and P. Kotyczka. Finite differences on staggered grids preserving the port-Hamiltonian structure with application to an acoustic duct. *Journal of Computational Physics*, 373:673–697, 2018.

[189] M. V. Trivedi, R. N. Banavar, and P. Kotyczka. Hamiltonian modelling and buckling analysis of a nonlinear flexible beam with actuation at the bottom. *Mathematical and Computer Modelling of Dynamical Systems*, 22(5):475–492, 2016.

[190] A. J. van der Schaft. Port-Hamiltonian differential-algebraic systems. In *Surveys in Differential-Algebraic Equations I*, pages 173–226. Springer, 2013.

[191] A. J. van der Schaft. *L2-Gain and Passivity Techniques in Nonlinear Control*. Springer, 3rd edition, 2017.

[192] A. J. van der Schaft, D. Jeltsema, et al. Port-Hamiltonian Systems Theory: An Introductory Overview. *Foundations and Trends in Systems and Control*, 1(2-3):173–378, 2014.

[193] A. J. van der Schaft and B. Maschke. The Hamiltonian formulation of energy conserving physical systems with external ports. *Archiv für Elektronik und Übertragungstechnik*, 49(5-6):362–371, 1995.

[194] A. J. van der Schaft and B. Maschke. A port-Hamiltonian formulation of open chemical reaction networks. In *Advances in the Theory of Control, Signals and Systems with Physical Modeling*, pages 339–348. Springer, 2010.

[195] A. J. van der Schaft and B. Maschke. Generalized port-Hamiltonian DAE systems. *Systems & Control Letters*, 121:31–37, 2018.

[196] A. J. van der Schaft and B. Maschke. Geometry of thermodynamic processes. *Entropy*, 20(12):925, 2018.

[197] A. J. van der Schaft and B. M. Maschke. Hamiltonian formulation of distributed-parameter systems with boundary energy flow. *Journal of Geometry and Physics*, 42(1):166–194, 2002.

[198] A. J. van der Schaft and B. M. Maschke. Discrete conservation laws and port-Hamiltonian systems on graphs and complexes. *arXiv:1107.2006v1*, 2011.

[199] A. J. van der Schaft and B. M. Maschke. Port-Hamiltonian systems on graphs. *SIAM Journal on Control and Optimization*, 51(2):906–937, 2013.

[200] J. Vankerschaver, H. Yoshimura, and J. E. Marsden. Multi-Dirac structures and Hamilton-Pontryagin principles for Lagrange-Dirac field theories. *arXiv preprint arXiv:1008.0252*, 2010.

[201] J. A. Villegas. *A port-Hamiltonian approach to distributed parameter systems*. PhD thesis, University of Twente, 2007.

[202] J. A. Villegas, H. Zwart, Y. Le Gorrec, B. Maschke, and A. J. van der Schaft. Stability and stabilization of a class of boundary control systems. In *44th IEEE Conference on Decision and Control and European Control Conference, Sevilla*, pages 3850–3855, 2005.

[203] G. Viola, R. Ortega, R. Banavar, J. A. Acosta, and A. Astolfi. Total energy shaping control of mechanical systems: Simplifying the matching equations via coordinate changes. *IEEE Transactions on Automatic Control*, 52(6):1093–1099, 2007.

[204] N. M. T. Vu, L. Lefèvre, and B. Maschke. A structured control model for the thermo-magneto-hydrodynamics of plasmas in Tokamaks. *Mathematical and Computer Modelling of Dynamical Systems*, 22(3):181–206, 2016.

[205] N. M. T. Vu, L. Lefèvre, R. Nouailletas, and S. Brémond. Symplectic spatial integration schemes for systems of balance equations. *Journal of Process Control*, 51:1–17, 2017.

[206] N. M. T. Vu, R. Nouailletas, L. Lefèvre, and F. Felici. Plasma q-profile control in Tokamaks using a damping assignment passivity-based approach. *Control Engineering Practice*, 54:34–45, 2016.

[207] M. Wang, A. Bestler, and P. Kotyczka. Modeling, discretization and motion control of a flexible beam in the port-Hamiltonian framework. *IFAC-PapersOnLine*, 50(1):6799–6806, 2017.

[208] K. F. Warnick and P. H. Russer. Differential forms and electromagnetic field theory. *Progress In Electromagnetics Research*, 148:83–112, 2014.

[209] H. Whitney. *Geometric Integration Theory*. Princeton University Press, 1957.

[210] F. Woittennek. On flatness and controllability of simple hyperbolic distributed parameter systems. In *18th IFAC World Congress, Milano, Italy*, pages 14452–14457, 2011.

[211] F. Woittennek and J. Rudolph. Controller canonical forms and flatness-based state feedback for 1D hyperbolic systems. *IFAC Proceedings Volumes*, 45(2):792–797, 2012.

[212] T. Wolf, B. Lohmann, R. Eid, and P. Kotyczka. Passivity and structure preserving order reduction of linear port-Hamiltonian systems using Krylov subspaces. *European Journal of Control*, 16(4):401–406, 2010.

[213] H. Yoshimura and J. E. Marsden. Dirac structures in Lagrangian mechanics Part II: Variational structures. *Journal of Geometry and Physics*, 57(1):209–250, 2006.

[214] W. Zhou, B. Hamroun, F. Couenne, and Y. Le Gorrec. Distributed port-Hamiltonian modelling for irreversible processes. *Mathematical and Computer Modelling of Dynamical Systems*, 23(1):3–22, 2017.

[215] H. Zwart, Y. Le Gorrec, and B. Maschke. Building systems from simple hyperbolic ones. *Systems & Control Letters*, 91:1–6, 2016.

[216] H. Zwart, Y. Le Gorrec, B. Maschke, and J. Villegas. Well-posedness and regularity of hyperbolic boundary control systems on a one-dimensional spatial domain. *ESAIM: Control, Optimisation and Calculus of Variations*, 16(4):1077–1093, 2010.