# Chapter 1

# Introduction

Traditionally, theory and experiment are the two classical bases of natural sciences such as physics. Additionally, computational methods have emerged, especially during the last decades. Depending on the context, they can be seen as enhancements of experiments and theory or even as a third pillar of natural sciences. Due to the power of modern computers, the applicability of numerical schemes has increased drastically. Today, computational methods have become important components of experiments (e.g. via data processing) and theory (e.g. via simulations). Since the theory can be so complex that analytical approaches to certain questions become intractable, numerical approaches are inevitable. Moreover, physical experiments can be too expensive or impossible, augmenting the necessity of numerical experiments. This thesis is dedicated to the investigation and development of numerical methods for hyperbolic partial differential equations arising in continuum physics and contains several deep and new theoretical and practical insights which have resulted in novel numerical algorithms that are provably stable and robust, presented here for the first time as a whole.

Exactly as physical experiments and theory, computational methods have to be studied carefully. If new theoretical insights should be gained or verified through numerical approaches, e.g. via the approximate solution of (partial) differential equations, it is necessary to trust these methods. While it is inevitable to verify the numerical schemes similar to new theories using carefully designed test cases, sanity checks with settled theories have to be performed as well.

In the following, balance laws relying on fundamental physical principles such as the conservation of mass, momentum, and energy will be investigated. If small-scale and viscous effects are neglected, many physical systems with finite speed of propagation can be described by hyperbolic balance laws. Some examples that will be studied in this thesis are the shallow water equations with varying bottom topography and the Euler equations of gas dynamics.

For linear equations, the mathematical theory is considerably well developed and can be used as a guideline for the construction of numerical schemes, enabling a very high level of trust in these computational methods. For general balance laws, the conservative form of the equations is an important aspect that should be handled correctly. Furthermore, well-posedness of linear equations can often be studied by the energy method, relying on integration by parts and its multidimensional analogues such as the divergence theorem and Green's identities.

Following the fundamental idea that computational methods should be based on the same grounds as the corresponding analytical theories, summation-by-parts (SBP) operators have emerged in the framework of finite difference methods for hyperbolic equations, cf. [71, 113, 114, 161, 271, 278]. Since these are designed to mimic integration by parts discretely, techniques used at the continuous level can be transferred to numerical schemes. Thus, proofs of well-posedness obtained by the energy method can be translated to proofs of energy stability at the semidiscrete level. In this way, many stability and conservation properties of numerical schemes can be obtained by mimicking analytical techniques used in the theory of linear hyperbolic equations.

However, the theory of general nonlinear hyperbolic balance laws is far less developed than the corresponding linear theory. Without the analytical guideline, the construction of reliable numerical methods might seem to be doomed to failure. Nevertheless, many analytical insights such as some fundamental existence results have been gained from numerical approaches, cf. [19, 100, 134, 166]. While there are many established results for scalar conservation laws and low order numerical methods as in [47], the situation is different for multidimensional systems and high order methods. However, it is still sensible to mimic properties at the continuous level discretely.

Since solutions of nonlinear hyperbolic balance laws typically develop discontinuities in finite time, weak solutions have to be considered. In order to obtain well-posedness for scalar problems, entropy conditions have to be imposed, cf. [49, 164, 165, 169, 211]. While there are many different approaches to such entropy conditions, a generalisation of the energy method to nonlinear hyperbolic equations will be considered in the following. Motivated by the second law of thermodynamics, balance laws of additional quantities such as the physical entropy for the Euler equations are considered. While entropy equalities are valid for sufficiently smooth solutions, entropy inequalities are postulated for physically relevant weak solutions.

Although it is unknown whether such entropy inequalities can be used to obtain well-posedness for general nonlinear hyperbolic conservation laws, they do provide some stability results for systems. Therefore, entropy stability has been investigated as a design principle of numerical schemes [284, 288]. While entropy stability alone is not sufficient to obtain convergence in general, it still seems to be a useful ingredient for computational methods. However, further robustness properties adapted to the specific problem have to be considered as well, such as the preservation of non-negativity of density and pressure for the Euler equations.

Therefore, it is useful to mimic important qualities of the continuous level discretely, relying on the basic idea of summation-by-parts operators. During the last years, there has been an enduring and increasing interest in generalised SBP operators and their application in various frameworks of numerical methods such as finite difference [71, 278], finite volume [205, 207], discontinuous Galerkin [89, 90, 92, 157, 215], and the recent flux reconstruction/correction procedure via reconstruction schemes [139, 140, 237, 297].

The main purpose of this thesis is to investigate such general summation-by-parts operators with respect to the construction of entropy stable numerical methods, possibly augmented with additional robustness and mimetic properties.

Roughly, two different aspects of stability can be considered: Stability of smooth solutions and stability of discontinuous solutions such as shock waves. While entropy conditions have been motivated above by problems occurring for non-smooth solutions, they provide some important stability results for smooth solutions as well.

The term entropy stability of numerical methods has been used to imply a (semi-) discrete entropy inequality, similarly to the continuous one. Relying on the philosophy to construct entropy conservative baseline schemes and add dissipation as necessary to stabilise the method, entropy stable schemes can be constructed, cf. [79, 284]. However, since they provide only the correct sign of the entropy dissipation but not necessarily the correct amount, they are in general not "stable", since oscillations and other problems can arise. Hence, "entropy dissipative" might be a better term than "entropy stable", although the latter term has been used traditionally and will therefore also be used in this thesis. Nevertheless, entropy stability can imply important stability aspects for smooth solutions such as in turbulence simulations.

Since entropy inequalities only provide the correct sign of the entropy dissipation, discontinuities pose additional challenges, especially for high order numerical schemes. In general, adapted techniques such as discontinuity detection, artificial dissipation, modal filtering, and the application of finite volume subcells seem to be necessary if high order methods are used to approximate discontinuous solutions. While this thesis is concerned with both aspects of stability, the investigations of high order methods using SBP operators focus more on entropy stability and less on explicit discontinuity handling.

This thesis is structured as follows. At first, a brief introduction to hyperbolic balance laws is given in chapter 2. Afterwards, chapter 3 provides a summary of fundamental concepts of numerical methods for hyperbolic balance laws. While there are many classical concepts, this chapter is focused especially on general summation-by-parts operators and recent results about entropy stability and high order methods using such general SBP operators. Here, new results obtained by the author are integrated into the presentation. Especially, high order of accuracy is proven for general flux difference discretisations (Theorem 3.16), extending the theory of [73, 170] significantly. Moreover, the basic technique to construct affordable, entropy conservative numerical fluxes is distilled into Procedure 3.28 and a general approach using several types of mean values is presented. Since the notations used in different communities vary significantly, translation rules are given in subsection 3.1.3, facilitating the reading of this thesis.

Based on these fundamental results, stability and conservation properties of numerical schemes using generalised SBP operators for some scalar equations are investigated in chapter 4. The construction of conservative and stable schemes using general summation-by-parts operators is enabled by the interpretation of these methods via

split forms of the underlying equations. In this way, entropy stability of numerical methods using general SBP operators has been obtained for the first time in the last years. This recent approach proposed by the author is described and extended to further conservation laws, cf. Theorem 4.3 and Proposition 4.27. Moreover, entropy stability in bounded domains is investigated and proven for adequate numerical fluxes, cf. Theorem 4.4. Since honest scientific work should include presentations of both advantages and disadvantages, limitations of this approach are also investigated in section 4.6.

Since there has been an increasing interest in general SBP operators and several shortcomings have been attributed to non-classical schemes in [187, 208], hyperbolic equations with varying coefficients are studied in chapter 5. Having introduced new ideas for general SBP operators in the previous chapter, extensions to these equations are developed. Although the investigated hyperbolic equations with varying coefficients are linear, they pose additional challenges for general SBP operators, since not all properties of the continuous level can be mimicked discretely as desired. However, it is still possible to construct stable and conservative schemes, cf. Theorems 5.4, 5.7, and 5.12. Moreover, these new approaches are compared to recent investigations by other researchers in sections 5.1.4, 5.2.3, and 5.3.3. Of course, extensive numerical results are also presented in section 5.4.

Turning to nonlinear systems of balance laws, the shallow water equations are investigated in chapter 6. In this recent work of the author, a general two-parameter family of entropy conservative numerical fluxes is investigated for the first time, cf. Lemma 6.2. Additionally, these fluxes are adapted to a varying bottom topography in a well-balanced way, enabling the preservation of non-trivial steady states, i.e. the lake-at-rest initial condition, cf. Lemma 6.12. Afterwards, the numerical fluxes are employed to create entropy stable high order discretisations using general SBP operators in Theorem 6.17. In order to ensure positivity preservation using the framework of Zhang and Shu [321], the entropy conservative fluxes are enhanced by dissipation operators and proven to be robust in the sense of positivity preservation under a non-vanishing CFL condition, cf. Lemma 6.24. Furthermore, finite volume subcells are applied to this setting (section 6.6) and several classical numerical fluxes are compared (sections 6.5 and 6.7).

Increasing the complexity of the systems, the Euler equations of gas dynamics are studied in chapter 7. To the author's knowledge, it is unknown whether stable and conservative methods of the same form as before can be constructed using general SBP operators, mainly due to the appearance of the logarithm in the physical (specific) entropy. Nevertheless, it is still possible to apply classical summation-by-parts operators. Presenting results published recently by the author, several numerical fluxes are developed in sections 7.2 and 7.3 and compared with existing ones. Using a semi-formalised notion of kinetic energy preservation, recent numerical experiments of [93] revealed that "kinetic energy preserving" numerical fluxes do not preserve the kinetic energy. Here, analytical insights into this behaviour are obtained in section 7.4,

especially in Theorems 7.2 and 7.3, resulting in a well-defined notion of kinetic energy preserving numerical fluxes (Definition 7.4). Using this new definition, a novel entropy conservative and kinetic energy preserving numerical flux is developed for the first time, cf. Theorem 7.8. Moreover, positivity preservation of entropy stable numerical fluxes is investigated and proven under appropriate conditions (Theorem 7.9), corroborating recent experimental results obtained in [55].

Turning to fully discrete stability of high order numerical schemes, chapter 8 focuses on time integration methods. In order to avoid the computationally expensive solution of implicit equations, explicit Runge-Kutta methods are investigated. A new procedure to prove stability of such schemes for linear equations with general semibounded operators is developed (Procedure 8.4) and applied to high order strong stability preserving methods, cf. Theorems 8.5, 8.7, and 8.10. However, such results seem to be limited to linear equations.

Thereafter, the main concept of the previous investigations is targeted again in chapter 9 — the entropy. Instead of using entropy stability as a design principle for numerical schemes, entropy variational ideas are investigated with respect to their applicability in computational methods for hyperbolic equations. Here, some ideas inspired by Dafermos [50, 51] are developed in section 9.1. Moreover, techniques from the classical ENO procedure are employed to construct new reconstruction schemes for finite volume methods in section 9.2. Additionally, new insights into some ENO methods are obtained by variational formulations, cf. Lemma 9.11 and Theorem 9.13. Extensive numerical investigations are conducted, contradicting the common folklore that numerical methods using polynomials with high degree are high order accurate. Therefore, extended studies and discussions are presented in section 9.3.

Finally, the results of this thesis are summed up in chapter 10. Furthermore, conclusions are drawn and some directions of further research are presented.

# Chapter 2

# Hyperbolic Balance Laws

In this chapter, some fundamental results concerning hyperbolic balance laws are summarised. Thorough introductions and results can be found in [17, 19, 49, 67, 101, 102, 114, 134, 165, 169, 172, 250, 251, 292, 300] and references cited therein.

## 2.1 Basic Definitions and Properties

Here, quasilinear equations of the form

$$\begin{aligned} \partial_t u(t,x) + \sum_{j=1}^{d} a^j\big(t,x,u(t,x)\big)\cdot\partial_j u(t,x) &= \tilde{s}\big(t,x,u(t,x)\big), && t\in(0,T),\ x\in\Omega,\\ u(0,x) &= u_0(x), && x\in\Omega, \end{aligned} \tag{2.1}$$

are considered, where $t$ denotes time, $x$ is the space coordinate, $\Omega\subset\mathbb{R}^d$ is an open set, and $u\colon [0,T)\times\Omega\to\Upsilon\subset\mathbb{R}^m$ is the sought solution. Additionally, the coefficients $a^j\colon (0,T)\times\Omega\times\Upsilon\to\mathbb{R}^{m\times m}$ and $\tilde{s}\colon (0,T)\times\Omega\times\Upsilon\to\mathbb{R}^m$ are given functions. The partial derivative with respect to time is denoted by $\partial_t$ and $\partial_j$ is the derivative with respect to the $j$-th space coordinate $x_j$. The initial datum $u_0$ is given to complement the Cauchy problem/*initial value problem* (IVP). Furthermore, boundary data can be considered if $\Omega\neq\mathbb{R}^d$, resulting in an *initial boundary value problem* (IBVP).

**Definition 2.1** (Hyperbolic equation). The quasilinear system (2.1) is called *hyperbolic*, if the matrix

$$\sum_{j=1}^{d} n_j a^j(t,x,y) \tag{2.2}$$

is diagonalisable over $\mathbb{R}$ for all $t\in(0,T)$, $x\in\Omega$, $y\in\Upsilon$, and $n=(n_1,\dots,n_d)^T\in\mathbb{S}^{d-1}$. The system is called *strictly hyperbolic*, if the eigenvalues $\lambda_1\le\dots\le\lambda_m$ of the matrix (2.2) are distinct. ◁

In the following, the partial derivative of a function $f^j\colon (0,T)\times\Omega\times\Upsilon\to\mathbb{R}^m$, $(t,x,y)\mapsto f^j(t,x,y)$, with respect to its third argument $y$ will be written as $\partial_u f^j$ instead of $\partial_y f^j$ since this notation is widely used.

**Definition 2.2** (Balance law). The quasilinear system (2.1) is called a *balance law*, if there exist *flux functions* $f^j\colon (0,T)\times\Omega\times\Upsilon\to\mathbb{R}^m$ such that $a^j(t,x,y)=\partial_u f^j(t,x,y)$, for all $j\in\{1,\dots,d\}$, $t\in(0,T)$, $x\in\Omega$, $y\in\Upsilon$. ◁

A balance law can be written in the general form

$$\begin{aligned}\partial_t u(t,x)+\partial_j f^j\big(t,x,u(t,x)\big)&=s\big(t,x,u(t,x)\big), && t\in(0,T),\ x\in\Omega,\\ u(0,x)&=u_0(x), && x\in\Omega,\end{aligned}\tag{2.3}$$

where summation over the repeated index $j$ is supposed. In general, the function $s\colon(0,T)\times\Omega\times\Upsilon\to\mathbb{R}^m$ is different from $\tilde{s}$ in (2.1) and is called *source term*. Here and in the following, the partial derivative operators $\partial_t$ and $\partial_j$ are assumed to act as linear operators on the terms at their right-hand side, i.e. $\partial_j f^j\big(t,x,u(t,x)\big)$ should be read as $\partial_j\Big(f^j\big(t,x,u(t,x)\big)\Big)$ instead of $\big(\partial_j f^j\big)\big(t,x,u(t,x)\big)$, where $t$,$x$, and $u(t,x)$ are inserted as arguments of the derivative $\partial_j f^j$.

If $u$ is a sufficiently smooth (e.g. continuously differentiable) solution of the balance law (2.3) and $\Omega_l\subset\Omega$ such that the divergence (Gauss) theorem can be applied,

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_l}u(t,x)\,\mathrm{d}x&=\int_{\Omega_l}\partial_t u(t,x)\,\mathrm{d}x\\ &=-\int_{\Omega_l}\partial_j f^j\big(t,x,u(t,x)\big)\,\mathrm{d}x+\int_{\Omega_l}s\big(t,x,u(t,x)\big)\,\mathrm{d}x\\ &=-\int_{\partial\Omega_l}n_j f^j\big(t,x,u(t,x)\big)\,\mathrm{dA}_x+\int_{\Omega_l}s\big(t,x,u(t,x)\big)\,\mathrm{d}x,\end{aligned}\tag{2.4}$$

where $n=(n_1,\dots,n_d)^T$ is the outer unit normal at $\partial\Omega_l$ and $\int_{\partial\Omega_l}\dots\mathrm{dA}_x$ is a surface integral. Thus, the rate of change of the integral of $u$ is given by the flux through the boundary and additional source terms, motivating the nomenclature "fluxes" and "source terms".

**Definition 2.3** (Conservation law). A balance law (2.3) is called a *conservation law*, if the source term $s\big(t,x,u(t,x)\big)$ vanishes. In this case, $u$ is also called a *conserved quantity* or *conserved variable*. ◁

Consequently, the rate of change of the integral of a conserved variable is given by the flux through the boundary of the given domain. Sometimes, $u$ is called conserved variable even in the case of general balance laws (2.3). In the rest of this work, mostly hyperbolic conservation laws are considered. In this chapter, the following examples in one space dimension without explicit dependence of the flux on time and space will be used.

**Example 2.4** (Linear advection with constant coefficients). The *linear advection equation with constant coefficients* is the conservation law

$$\begin{aligned}\partial_t u(t,x)+\partial_x\big(a\,u(t,x)\big)&=0,\\ u(0,x)&=u_0(x),\end{aligned}\tag{2.5}$$

possibly supplemented with suitable boundary conditions. Here, $a \in \mathbb{R}$ is the constant advection velocity and $u\colon [0,T)\times\mathbb{R}\to\mathbb{R}$ is a scalar valued function. ◁

**Example 2.5** (Burgers' equation). A nonlinear scalar conservation law is given by *Burgers' equation*

$$\begin{aligned} \partial_t u(t,x) + \partial_x \frac{1}{2} u(t,x)^2 &= 0, \\ u(0,x) &= u_0(x), \end{aligned} \tag{2.6}$$

again supplemented with boundary conditions if appropriate. ◁

In the theory of hyperbolic balance laws, several notions of solutions have been developed. For continuously differentiable functions $u$, they coincide with the following canonical solution concept.

**Definition 2.6** (Classical solution). A function $u \in C^1\big([0,T)\times\Omega;\Upsilon\big)$ is called a *classical solution* of the balance law (2.3) if it satisfies the equation pointwise. ◁

However, classical solutions will in general exist only locally in time. Indeed, consider a classical solution of the scalar conservation law

$$\partial_t u + \partial_x f(u) = 0, \qquad u(0) = u_0, \tag{2.7}$$

with smooth flux $f$ in one space dimension. A *characteristic* is a function $x\colon [0,T)\to\mathbb{R}$ satisfying $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = f'\big(u(t,x(t))\big)$ and $x(0) = x_0$. Thus,

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}u(t,x(t)) &= (\partial_t u)(t,x(t)) + (\partial_x u)(t,x(t))\ \tfrac{\mathrm{d}}{\mathrm{d}t}x(t) \\ &= -\big(\partial_x (f\circ u)\big)(t,x(t)) + (\partial_x u)(t,x(t))\, f'\big(u(t,x(t))\big) = 0, \end{aligned} \tag{2.8}$$

where $(\partial_t u)(t,x(t))$ is the value of the derivative $\partial_t u$ at $(t,x(t))$. Therefore, a classical solution $u$ is constant along characteristics, i.e.

$$\forall t\colon \quad u(t,x(t)) = u(0,x(0)) = u_0(x_0), \tag{2.9}$$

and the slope $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = f'\big(u(t,x(t))\big) = f'\big(u_0(x_0)\big)$ of characteristics is constant. Hence, $x(t) = x_0 + t f'(u_0(x_0))$, and a classical solution $u$ is given implicitly by

$$u(t,x) = u_0\big(x - t f'(u(t,x))\big). \tag{2.10}$$

However, since the slope of characteristics depends on the initial condition $u_0$, two characteristics starting at different origins can cross if $f'$ is not constant. Indeed, differentiating (2.10) yields

$$\begin{aligned} &\partial_x u(t,x) = u_0'\big(x - t f'(u(t,x))\big)\,\big(1 - t f''(u(t,x))\partial_x u(t,x)\big)\,, \\ \Longrightarrow\ &\partial_x u(t,x) = \frac{u_0'\big(x - t f'(u(t,x))\big)}{1 + t\, f''\Big(u_0\big(x - t f'(u(t,x))\big)\Big)\, u_0'\big(x - t f'(u(t,x))\big)}, \end{aligned} \tag{2.11}$$

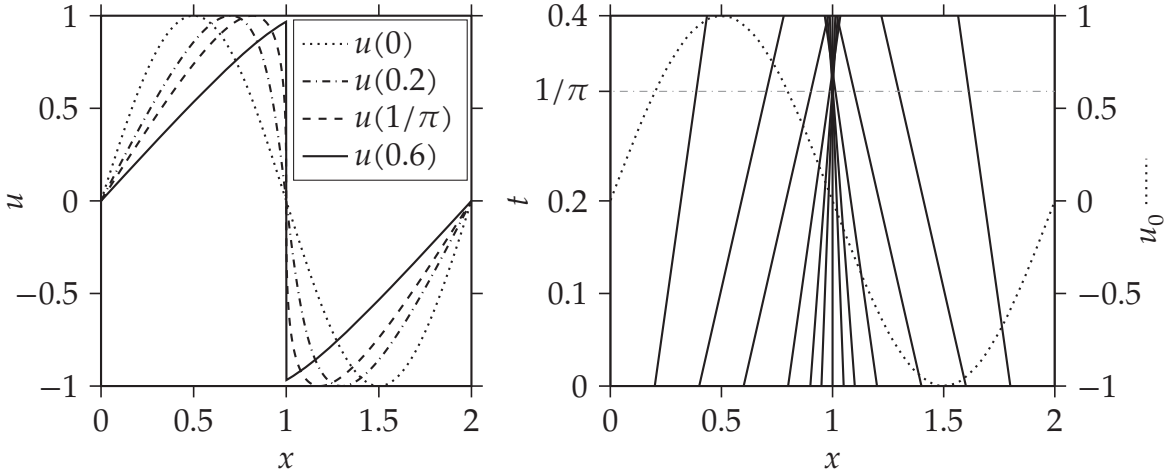
if the denominator is not zero. Here, (2.10) has been inserted in the second equation. Thus, depending on the flux $f$ and the slope $u_0'$ of the initial data, the derivative $\partial_x u(t,x)$ can blow up in finite time due to the development of discontinuities, as shown in the following example.

**Example 2.7** (Characteristics for Burgers' equation). Consider Burgers' equation (2.6) on the whole real line with initial condition $u_0(x) = \sin(\pi x)$. Since $f(u) = \frac{u^2}{2}$, the implicit equation (2.10) becomes $u(t,x) = u_0\big(x - t\,u(t,x)\big)$ and the derivative fulfils

$$\partial_x u(t,x) = \frac{u_0'\big(x - t\,u(t,x)\big)}{1 + t u_0'\big(x - t\,u(t,x)\big)}, \tag{2.12}$$

due to (2.11). If the slope $u_0'$ of the initial datum $u_0$ was non-negative, no blow-up of the derivative $\partial_x u$ can be expected for positive $t$. However, the initial datum $x \mapsto u_0(x) = \sin(\pi x)$ has also negative slopes with $\min(u_0') = -1/\pi$. Thus, seeking the smallest positive time $t$ for which the derivative $\partial_x u(t,x)$ blows up at some $x \in \mathbb{R}$ results due to (2.12) in $t = -1/\min(u_0') = 1/\pi$.

The solution $u$ obtained from (2.10) and some characteristics are visualised in Figure 2.1. As can be seen there, the characteristics starting at $x_0 \in (0,1)$ have a positive slope in the $(x,t)$ plane and the characteristics stating at $x_0 \in (1,2)$ have a negative slope. Thus, the values of the initial datum $u_0$ are transported from left to right in $(0,1)$ and from right to left in $(1,2)$. Hence, the slope of the solution $u(t)$ at $x = 1$ becomes steeper and a discontinuity develops at $t = 1/\pi$. Thereafter, a discontinuous solution $u$ can still be obtained by solving the implicit equation (2.10). ◁




(a) Solution $u(t,\cdot)$ at different times. (b) Characteristics starting at $t = 0$.

Figure 2.1: Visualisation of characteristics for Burgers' equation (2.6) with initial condition $u_0(x) = \sin(\pi x)$.