ON THE ACTIVE FLUX SCHEME FOR HYPERBOLIC PDES WITH SOURCE TERMS*

WASILIJ BARSUKOW†, JONAS P. BERBERICH‡, AND CHRISTIAN KLINGENBERG‡

Abstract. The active flux scheme is a finite volume scheme with additional point values distributed along the cell boundary. It is third order accurate and does not require a Riemann solver: the initial value problem at the particular points is solved instead. The intercell flux is then obtained from the evolved values along the cell boundary by quadrature. This paper focuses on the conceptual extension of active flux to include source terms, and thus for simplicity assumes the homogeneous part of the equations linear. To a large part the treatment of the source terms is independent of the choice of the homogeneous part of the system. Additionally, only systems are considered which admit characteristics (instead of characteristic cones). This is the case for scalar equations in any number of spatial dimensions and systems in one spatial dimension. Here, we succeed to extend the active flux method to include (possibly nonlinear) source terms while maintaining third order accuracy of the method. This requires a novel (approximate) operator for the evolution of point values and a modified update procedure of the cell average. For linear acoustics with gravity, it is shown how to achieve a well-balanced / stationarity preserving numerical method.

Key words. finite volume methods, active flux, source terms, balance laws, well-balanced methods, gravity

AMS subject classifications. 35L65, 35L45, 65M08

1. Introduction.

Numerous phenomena of the physical world are modeled by hyperbolic balance laws (conservation laws augmented by source terms). This includes gas dynamics, the motion of water waves, plasma physics and even general relativity. Often physical modeling requires to include source terms, and conservation is modified due to creation or annihilation of some of the evolved quantities. Chemical reactions, for example, change the number density of a species and produce or absorb heat (i.e. internal energy). Gravity accelerates matter downwards and creates momentum. In the shallow water model describing the motion of a free water surface the bottom topography enters the equations through a source term. Rewriting the hydrodynamic equations in a different coordinate system (e.g. in polar coordinates) makes geometric source terms appear. All these applications require reliable numerical methods which are able to deal with source terms.

Reliable numerical methods for hyperbolic conservation laws with source terms first need to perform well in the homogeneous case. This means for example that they need to cope with discontinuities / weak solutions and with phenomena arising in multiple spatial dimensions, such as involutions and non-trivial stationary states. This requirement has led [ER13, FR15] to suggest *active flux*, an extension of the finite volume method. Additionally to the cell average, this scheme evolves point values located at the cell boundary. The update of the point values is achieved by using an evolution operator that includes multi-dimensional information. The presence of the point values along the cell boundary then allows to compute the intercell flux

*Submitted to the editors DATE.

Funding: WB was supported by the German Academic Exchange Service (DAAD) with funds from the German Federal Ministry of Education and Research (BMBF) and the European Union (FP7-PEOPLE-2013-COFUND – grant agreement no. 605728) as well as by the Deutsche Forschungsgemeinschaft (DFG) through project 429491391 (BA 6878/1-1).

†Institute of Mathematics, Zurich University, 8057 Zurich, Switzerland (wasilij.barsukow@math.uzh.ch).

‡Wuerzburg University, Emil-Fischer-Strasse 40, 97074 Wuerzburg, Germany.

via quadrature. It has been shown in [BHKR19] that this scheme is stationarity preserving and vorticity preserving for linear acoustics without any fix. It is third order accurate. Extensions to nonlinear systems have been recently suggested e.g. in [Fan17, HKS19, Bar19a]. Active flux therefore seems to be promising for resolving many of the structure preservation problems that currently available methods are facing (an overview of existing methods for balance laws is given below).

In view of the many applications that involve source terms, this paper therefore aims at deriving the necessary modifications for active flux to be applicable to balance laws while retaining its third order accuracy. Including the source term requires a number of modifications. The homogeneous part of the equations therefore is for simplicity assumed to be a linear hyperbolic system for which characteristics are available. This is the case for scalar equations in any number of spatial dimensions and for systems in one spatial dimension. For multi-dimensional systems, the concept of characteristics needs to be replaced by characteristics cones. In the homogeneous case, active flux has been used for this situation as well ([ER13, BHKR19]), but an extension to inhomogeneous systems in multi-d, and to nonlinear systems remains subject of future work. To a large part, the strategies presented in this paper will, however, remain valid when the homogeneous part of the equations is nonlinear as well, and even for nonlinear multi-dimensional systems.

As soon as a source term is added to a hyperbolic system, new stationary states arise which often are of particular interest. The stationarity is due to the flux divergence being equal to the source term. Many areas of application of balance laws involve studies of dynamics on top of such an equilibrium (e.g. astrophysics, meteorology, tsunami modeling, . . .). This requires the numerical method to be very accurate on the stationary states in order to avoid spurious, artificial perturbations. Therefore the error of a numerical solution representing one of those stationary states should not increase with time, thus allowing the simulation to run for a long time.

Numerical methods which achieve this are called *well-balanced*, introduced in [GL96]. They make sure that the discretization of the flux divergence and the discretization of the source term match, and that the numerical method keeps the desired stationary state exactly stationary for any resolution of the grid. The concept of well-balanced methods has been extensively used in the context of shallow water equations with non-flat bottom topography (e.g. [ABB+04, BV94, LeV98] and references therein). Here, the balance is the so-called lake-at-rest solution, which amounts to an algebraic condition and can thus be given explicitly.

Another area in which well-balanced methods have high relevance is the simulation of hydrodynamic processes using compressible Euler equations with gravitational source term. The so-called hydrostatic state (stationary state with no velocity) is described by one PDE for two unknown functions. There are many hydrostatic states, depending on the additional thermodynamical relation that one chooses in order to close this PDE. The fact that the stationary state is itself given by a differential equation that cannot be integrated makes well-balancing much more delicate in this context. There are two different ways which are currently used to construct well-balanced methods for the Euler equations with gravity. The first and more traditional way is to restrict the class of hydrostatic solutions which are balanced exactly or to choose a particular, but arbitrary hydrostatic state (e.g. [CL94, LGB11, DZBK16, CK15, BCK16, CCK+18, BCKR19, BCK19]). This is advantageous in all those applications where the stationary state is known, and the evolution of perturbations around it shall be studied. If no information on the stationary state can be assumed, then the only way to proceed is to make sure that the stationary states of the numerical method are fulfilling some *discretization* of the corresponding PDE (e.g. [DZBK14, KM16, BKCK20]).

For linear numerical methods a theory of such *stationarity preserving* methods was given in [Bar19b], with a particular emphasis on this latter, more complicated, situation of the stationary states given by PDEs, and not by algebraic relations. It turns out that many standard numerical methods add diffusion even to those states that should remain stationary. The set of states that are actually kept stationary by such methods is very small (e.g. uniform constants). Stationarity preserving methods do not apply diffusion to certain discrete data. These data are described by a discrete version of the PDE governing the stationary states. Stationarity preserving methods thus keep stationary a much larger set of initial data. Independently of how these discrete equations actually look like, it is their existence that makes a qualitative difference. In a non-stationarity-preserving method, initial data sampled from an analytic stationary state will decay due to the diffusion and become unrecognizable in the end. In a stationarity preserving method, these initial data will evolve towards one of the many discrete stationary states approximating the steady PDE, and will remain there forever (up to machine precision). The long-time numerical solution will then indeed approximate the analytic stationary state. For more details, see [Bar19b]. In this paper we understand the concept of well-balancing in this sense of stationarity preservation.

In this paper, after extending the active flux scheme to include source terms, we construct a well-balanced active flux method for the equations of acoustics with gravity. The hydrostatic solutions of acoustics with gravity are comparable to those of the compressible Euler equations with gravity, since they are given via the same underdetermined differential equation. We show that the active flux scheme endowed with an exact evolution operator is intrinsically well-balanced in this way. In practice, an approximate evolution operator needs to be used. Hence we introduce a modification of the approximate evolution operator which makes the scheme well-balanced even upon usage of an approximate evolution operator.

The paper is organized as follows: After the active flux scheme for homogeneous problems is introduced in section 2, the modifications necessary for including source terms are discussed. Section 3 discusses the evolution operators necessary for the update of the point values. Section 4 is devoted to the modifications in the update of the average. Here, the focus lies on linear systems of equations with possibly nonlinear source terms in one spatial dimension and on linear advection in multiple spatial dimensions. Section 5 discusses well-balancing of active flux for linear acoustics with gravity. Section 6 finally demonstrates numerically that the new method attains third order accuracy with linear and nonlinear source terms, can be used to compute Riemann problems, and displays well-balanced behavior for stationary states.

This work can be seen in the larger context of the quest for structure preserving numerical methods, of which well-balanced methods form an example. Extending these results to nonlinear hyperbolic equations with source terms and thus combining the structure preserving properties of active flux remains subject of future work. However, the procedures suggested in this paper are formulated with as little reference to the linearity of the equations as possible.

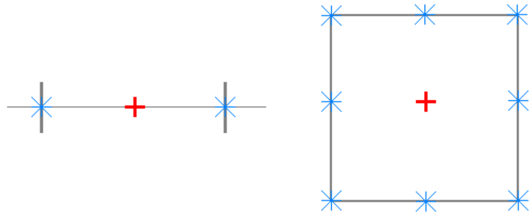

FIG. 1. *The degrees of freedom used for active flux. Stars indicate the location of point values, and the cross (placed in the center symbolically) refers to the cell average.* Left*: One spatial dimension.* Right*: Two spatial dimensions.*

2. The active flux scheme. Consider the initial value problem for an $m \times m$ system of hyperbolic balance laws in d spatial dimensions[1]

$$\partial_t q + \nabla \cdot \mathbf{f}(q) = s(q) \qquad q : \mathbb{R}_0^+ \times \mathbb{R}^d \to \mathbb{R}^m,\ f, s : \mathbb{R}^m \to \mathbb{R}^m \tag{2.1}$$

$$q(0, \mathbf{x}) := q_0(\mathbf{x}) \tag{2.2}$$

This section reviews the general idea of the active flux scheme. Some of the details then depend on the particular equation that is to be solved. After the general concept is outlined, the details that make it applicable to hyperbolic balance laws are discussed in sections 3 and 4.

2.1. Degrees of freedom in the active flux scheme. The active flux scheme ([ER13, BHKR19], first introduced in [VL77]) is an extension of the finite volume scheme. The active flux scheme evolves both the cell average and point values which are distributed along the cell boundary. In particular, here the following two choices are considered (see Figure 1):

- In one spatial dimension, there is a point value $q_{i+\frac{1}{2}}$ located at each cell interface $x_{i+\frac{1}{2}}$. Thus every cell has access to one cell average $\bar{q}_i$ and two point values at its interfaces.
- On Cartesian grids in two spatial dimensions, there is a point value $q_{i+\frac{1}{2},j}$, $q_{i,j+\frac{1}{2}}$ at each edge midpoint and one at each node $q_{i+\frac{1}{2},j+\frac{1}{2}}$. Every cell has access to one cell average $\bar{q}_{ij}$ and 8 point values distributed along the cell interface.

Note that the point values at cell interfaces are shared by the adjacent cells. Thus, in one spatial dimension, on average there are 2 degrees of freedom per cell: 1 cell average and 2 interface values shared each by 2 cells. In two spatial dimensions in the setup as described above there are 4 degrees of freedom per cell: 1 cell average, 4 edge values, each shared by two cells and 4 node values each shared by 4 cells.

Note also that active flux does not use a staggered grid. The degrees of freedom at the cell boundaries are not averages over staggered volumes, but point values. This also explains why there is no notion of a conservative update for these, because this concept only applies to averages. The update of the cell average in the active flux method is, of course, conservative (see below).

2.2. Update of the cell average. As the active flux scheme is an extension of the finite volume scheme, given a numerical flux, the update of the average happens in

[1]In this paper, indices never denote derivatives. Boldface symbols denote vectors that have the same dimension as the space.

the same way as for finite volume schemes. In this section, this finite volume aspect of active flux is described in an arbitrary number of spatial dimensions. The numerical flux, however, is obtained very differently in the active flux scheme ([ER13, FR15]). This is then described in detail in section 4.

Consider the computational domain to be subdivided into polygonal computational cells. Upon integration of (2.1) over one time step $[t^n, t^n + \Delta t]$ and over one computational cell $\mathcal{C}$ one obtains an evolution equation for the cell average $\bar{q}_\mathcal{C} := \frac{1}{|\mathcal{C}|}\int_\mathcal{C} \mathrm{d}\mathbf{x}\, q(t, \mathbf{x})$:

$$\frac{\bar{q}_\mathcal{C}^{n+1} - \bar{q}_\mathcal{C}^n}{\Delta t} + \frac{1}{|\mathcal{C}|}\frac{1}{\Delta t}\int\limits_{t^n}^{t^n+\Delta t} \mathrm{d}t \int_{\partial\mathcal{C}} \mathrm{d}\sigma\, \mathbf{n}\cdot\mathbf{f}(q(t,\mathbf{x})) = \frac{1}{\Delta t}\int\limits_{t^n}^{t^n+\Delta t} \mathrm{d}t \frac{1}{|\mathcal{C}|}\int_\mathcal{C} \mathrm{d}\mathbf{x}\, s(q(t,\mathbf{x}))$$

Here, as usual, the index of the time step is denoted as a superscript and $q_\mathcal{C}^n$ denotes the average in cell $\mathcal{C}$ at time t^n. The boundary $\partial\mathcal{C}$ consists of edges e, such that one can rewrite

$$\frac{\bar{q}_\mathcal{C}^{n+1} - \bar{q}_\mathcal{C}^n}{\Delta t} + \frac{1}{|\mathcal{C}|}\frac{1}{\Delta t}\int\limits_{t^n}^{t^n+\Delta t} \mathrm{d}t \sum_{e\subset\partial\mathcal{C}}\int_e \mathrm{d}\sigma\, \mathbf{n}_e\cdot\mathbf{f}(q(t,\mathbf{x})) = \frac{1}{\Delta t}\int\limits_{t^n}^{t^n+\Delta t} \mathrm{d}t \frac{1}{|\mathcal{C}|}\int_\mathcal{C} \mathrm{d}\mathbf{x}\, s(q(t,\mathbf{x}))$$

The vector $\mathbf{n}_e$ is the outward unit normal of edge e. This expression, so far exact, becomes a finite volume scheme upon replacing the exact normal flux and source averages by suitable approximations $\hat{f}_e$ and $\hat{s}_\mathcal{C}$:

$$\frac{\bar{q}_\mathcal{C}^{n+1} - \bar{q}_\mathcal{C}^n}{\Delta t} + \frac{1}{|\mathcal{C}|}\sum_{e\subset\partial\mathcal{C}} |e|\hat{f}_e = \hat{s}_\mathcal{C} \tag{2.3}$$

with

$$\hat{f}_e \simeq \frac{1}{\Delta t}\int\limits_{t^n}^{t^n+\Delta t} \mathrm{d}t\, \frac{1}{|e|}\int_e \mathrm{d}\sigma\, \mathbf{n}_e\cdot\mathbf{f}(q(t,\mathbf{x})) \tag{2.4}$$

$$\hat{s}_\mathcal{C} \simeq \frac{1}{\Delta t}\int\limits_{t^n}^{t^n+\Delta t} \mathrm{d}t\, \frac{1}{|\mathcal{C}|}\int_\mathcal{C} \mathrm{d}\mathbf{x}\, s(q(t,\mathbf{x})) \tag{2.5}$$

Usual finite volume schemes introduce a (piecewise continuous) reconstruction of the averages, and obtain the numerical flux by an exact or approximate short-time evolution of this reconstruction. For example, introducing a piecewise constant function whose averages match the given cell averages, and solving the Riemann problems at the cell interfaces allows to compute a numerical flux.

The active flux scheme does not need this. Indeed, the point values along the boundary can be used to immediately approximate (2.4)–(2.5) by quadrature. The desired properties (most importantly the desired order of accuracy) of the resulting scheme dictate the number of point values along each edge and also the points in time at which these point values need to be available.

The source term also contributes to the update of the cell average. The quadrature necessary to approximate the source term average (2.5) to sufficient order in space and time is suggested in this paper for the first time and discussed in section 4.

2.3. Update of the point values. The cell average update, and in particular the computation of the intercell fluxes, requires accurate point values at the cell boundary to be available.

First consider the case where the source term vanishes: $s = 0$. For third order of accuracy, the integrals in (2.4) need to be approximated by Simpson's rule. For the integration in space this can easily be achieved using the available point values at each cell interface as described in section 2.1. For the integration in time all point values need to be available at $t^n, t^n + \frac{\Delta t}{2}$ and $t^n + \Delta t$. Altogether this yields a space-time Simpson rule.

In order to obtain sufficiently accurate time evolved point values, in [VL77] it has been suggested to reconstruct the data and to use an exact evolution operator. An exact evolution operator generally is unavailable for nonlinear problems, and therefore in [Fan17, HKS19, Bar19a] approximate evolution operators have been proposed. Even for linear systems of hyperbolic balance laws it is generally very difficult to obtain closed-form exact evolution operators, as is shown in section 3.2. Therefore the point values in the active flux scheme shall be evolved using a sufficiently high order *approximate* evolution operator applied to a reconstruction of the discrete data. An exact evolution operator provides the necessary upwinding in order to guarantee stability, and an approximate evolution operator needs to do the same. The approximate evolution operator is introduced in section 3.3.

2.4. Reconstruction. The reconstruction shall interpolate the point values and its average over the computational cell shall match the given cell average. In the following, to simplify notation, in one spatial dimension a uniform grid is assumed, although the reconstruction can immediately be generalized to nonuniform grids. In two spatial dimensions, a Cartesian grid is used. As mentioned in section 2.1, in one spatial dimension every cell has access to 3 degrees of freedom which makes a parabolic reconstruction natural. With the above-mentioned setup it is unique and reads ([VL77, FR15])

$$q_{\mathrm{recon},i}(x) = -3(2\bar{q}_i - q_{i-\frac{1}{2}} - q_{i+\frac{1}{2}})\frac{(x-x_i)^2}{\Delta x^2} \tag{2.6}$$

$$+ (q_{i+\frac{1}{2}} - q_{i-\frac{1}{2}})\frac{x-x_i}{\Delta x} + \frac{6\bar{q}_i - q_{i-\frac{1}{2}} - q_{i+\frac{1}{2}}}{4} \qquad x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \tag{2.7}$$

In two spatial dimensions as described above, every cell has access to 9 degrees of freedom, and there is a unique biparabolic reconstruction, which reads

$$
\begin{aligned}
q_{\mathrm{recon},ij}(\xi\Delta x,\eta\Delta y) := {} & \frac{9}{4}\bar{q}_{ij}\left(-1+4\xi^2\right)\left(-1+4\eta^2\right) \\
& -\frac{1}{4}q_{\mathrm{W}}\left(-1-4\xi+12\xi^2\right)\left(-1+4\eta^2\right) \\
& -\frac{1}{4}q_{\mathrm{E}}\left(-1+4\xi+12\xi^2\right)\left(-1+4\eta^2\right) \\
& -\frac{1}{4}q_{\mathrm{S}}\left(-1+4\xi^2\right)\left(-1-4\eta+12\eta^2\right) \\
& -\frac{1}{4}q_{\mathrm{N}}\left(-1+4\xi^2\right)\left(-1+4\eta+12\eta^2\right) \\
& +\frac{1}{16}q_{\mathrm{SW}}(-1+2\xi)(-1+2\eta)(-1-2\eta+2\xi(-1+6\eta)) \\
& +\frac{1}{16}q_{\mathrm{SE}}(1+2\xi)(-1+2\eta)(1+2\eta+2\xi(-1+6\eta)) \\
& +\frac{1}{16}q_{\mathrm{NW}}(-1+2\xi)(1+2\eta)(1-2\eta+2\xi(1+6\eta)) \\
& +\frac{1}{16}q_{\mathrm{NE}}(1+2\xi)(1+2\eta)(-1+2\eta+2\xi(1+6\eta))
\end{aligned} \tag{2.8}
$$

with $\xi := x/\Delta x$, $\eta := y/\Delta y$ and

$$q_{\mathrm{NE}} = q_{i+\frac{1}{2},j+\frac{1}{2}} \qquad q_{\mathrm{NW}} = q_{i-\frac{1}{2},j+\frac{1}{2}} \qquad q_{\mathrm{SW}} = q_{i-\frac{1}{2},j-\frac{1}{2}} \qquad q_{\mathrm{SE}} = q_{i+\frac{1}{2},j-\frac{1}{2}} \tag{2.9}$$

$$q_{\mathrm{N}} = q_{i,j+\frac{1}{2}} \qquad q_{\mathrm{S}} = q_{i,j-\frac{1}{2}} \qquad q_{\mathrm{E}} = q_{i+\frac{1}{2},j} \qquad q_{\mathrm{W}} = q_{i-\frac{1}{2},j} \tag{2.10}$$

Note that both reconstructions are globally continuous, but generally not continuously differentiable at the cell interfaces.

2.5. Overview of the algorithm. The overall algorithm of active flux is as follows:

1. Given cell averages and point values, compute a reconstruction according to section 2.4.
2. Use the reconstruction as initial data in the update of the point values. The choices of evolution operators considered so far are discussed in section 2.3 and evolution operators in presence of source terms are suggested in section 3.3 below.
3. Given the updated point values along the cell interfaces, compute the intercell fluxes via quadrature (sections 2.2 and 4 for the homogeneous and the inhomogeneous cases, respectively).
4. Update the cell averages via (2.3).

A CFL-type condition arises in the update of the point values: the domain of dependence of the evolution operator needs to be contained in the neighbouring cells. Denoting by $\lambda_{\max}$ the maximum speed of propagation, the time step needs to be chosen as

$$\Delta t \leq \frac{L_{\min}}{\lambda_{\max}} \tag{2.11}$$

where $L_{\min} = \Delta x$ in one spatial dimension, and $L_{\min} = \frac{1}{2}\min(\Delta x, \Delta y)$ in two spatial dimensions, when the point values are distributed as described in section 2.1.

3. Evolution of the point values in presence of a source term. The evolution of the point values needs to account for the source term. Additionally, in this paper a special focus shall lie on structure preservation properties of the resulting scheme. In the homogeneous case such properties have been observed upon usage of an exact evolution operator ([BHKR19]). In presence of a source term, one needs to use an approximate evolution operator (section 3.3), but should nevertheless aim at making it such that it does not spoil structure preservation (see section 5).

For certain equations, the inhomogeneous problem admits an exact solution (sections 3.1–3.2). This is valuable in order to assess specific properties of the numerical method later.

3.1. Linear advection with a source term in multiple spatial dimensions. Consider a scalar equation ($m = 1$) and $\mathbf{f}(q) = \mathbf{U}q$ with $\mathbf{U} \in \mathbb{R}^d$. Then

$$\partial_t q + \mathbf{U} \cdot \nabla q = s(q) \tag{3.1}$$

amounts to the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t} q = s(q) \tag{3.2}$$

along the straight characteristic of velocity $\mathbf{U}$. This ODE can be easily solved analytically:

$$\int_{q_0(\mathbf{x}-\mathbf{U}t)}^{q(t,\mathbf{x})} \frac{\mathrm{d}p}{s(p)} = t \tag{3.3}$$

E.g. for $s(q) = \kappa q$ this yields $\ln \frac{q(t,\mathbf{x})}{q_0(\mathbf{x}-\mathbf{U}t)} = \kappa t$, or

$$q(t, \mathbf{x}) = q_0(\mathbf{x} - \mathbf{U}t) \exp(\kappa t) \tag{3.4}$$

and for $s(q) = \kappa q^B$, $B \neq 1$

$$q(t, \mathbf{x}) = \Big((q_0(\mathbf{x} - \mathbf{U}t))^{1-B} + (1 - B)\kappa t \Big)^{\frac{1}{1-B}} \tag{3.5}$$

3.2. Linear acoustics with gravity in one spatial dimension. This section has threefold purpose. First, it introduces the acoustic equations with a gravity source term, which form a very useful system for the study of structure preservation of numerical methods. This is the set of equations for which a well-balanced method is derived in 5. This section also demonstrates the difficulties of finding an exact solution to an inhomogeneous system even if it is linear. Finally, the exact solution derived here is used later in order to assess the accuracy of the numerical method.

The equations of linear acoustics in one spatial dimension endowed with a gravity source term read:

$$\partial_t \rho + \partial_x v = 0 \tag{3.6}$$
$$\partial_t v + \partial_x p = \rho g \qquad g \in \mathbb{R} \tag{3.7}$$
$$\partial_t p + c^2 \partial_x v = 0 \tag{3.8}$$

The corresponding homogeneous problem (linear acoustics) is the linearization of the Euler equations around the background state of constant density $\rho_{\mathrm{bg}} = 1$, constant pressure $p_{\rm bg}$ and vanishing velocity. Then the speed of sound $c = \sqrt{\frac{\gamma p_{\rm bg}}{\rho_{\rm bg}}}$ is a constant ($\mathbb{R} \ni \gamma > 1$). The full system (3.6)–(3.8) can be understood as a particular kind of a linearization of the Euler equations with gravity[2]

$$\partial_t \rho + \partial_x(\rho v) = 0 \tag{3.9}$$

$$\partial_t(\rho v) + \partial_x(\rho v^2 + p) = \rho g \tag{3.10}$$

$$\partial_t e + \partial_x(v(e+p)) = 0 \tag{3.11}$$

$$e = \frac{p}{\gamma - 1} + \frac{1}{2}\rho v^2 - \rho g x \tag{3.12}$$

The static (stationary and $v = 0$) states of (3.9)–(3.11) are governed by $\partial_x p = \rho g$. This equation can only be solved if e.g. ρ is given as a function of x, or if another relation is provided between any two of the variables p, ρ, e. This multitude of possible stationary states is reflected in the linearization (3.6)–(3.8). (This is the reason for this particular choice of a linearization.) Observe that stationary states of (3.6)–(3.8) also are governed by $\partial_x p = \rho g$ and that p can only be computed if ρ is given as a function of x, or if an additional relation is provided that links ρ and p. This is an example of a so-called non-trivial stationary state as introduced in [Bar19b]. Examples of stationarity preserving schemes for (3.6)–(3.8) have been discussed in [Bar18].

The exact solution of (3.6)–(3.8) is studied in the Appendix A. This solution is not part of the suggested method but only serves auxiliary purposes, such as accuracy checks. However it illustrates the difficulties encountered when solving linear systems with sources. To the authors' knowledge the exact solution to (3.6)–(3.8) is not available in the literature so far.

3.3. Runge-Kutta method for linear systems with a source. Consider an $m \times m$ linear system in characteristic variables:

$$(\partial_t + \lambda_\ell \partial_x) Q_\ell = S_\ell(Q_1, \dots, Q_m) \qquad \ell = 1, \dots, m \tag{3.13}$$

From now on, the capital letter Q denotes the characteristic variables of this particular system, whereas q continues to denote a generic variable.

Recall the following theorem from [Bar19a]:

THEOREM 3.1. *Assume a hyperbolic CFL condition $\Delta x/\Delta t \to \text{const}$ as $\Delta t \to 0$. If the approximate evolution $Q^{\rm approx}(t,x)$ approximates the exact solution $Q(t,x)$ for fixed x at least as*

$$Q^{\rm approx}(t,x) = Q(t,x) + \mathcal{O}(t^3) \tag{3.14}$$

and the quadrature rules used to approximate (2.4)–(2.5) yield the exact value up to an error of $\mathcal{O}(\Delta t^\alpha \Delta x^\beta)$, $\alpha + \beta \geq 3$ then active flux formally achieves third order accuracy.

Note that the simple approach of evolving each component of the source term along its associated characteristic

$$Q_\ell(t,x) \simeq Q_{\ell,0}(x - \lambda_\ell t) + t S_\ell(Q_{1,0}(x - \lambda_\ell t), \dots, Q_{m,0}(x - \lambda_\ell t)) \qquad \ell = 1, \dots, m \tag{3.15}$$

[2] Note that often the energy equation is written with a source term $\rho g v$ appearing. This source term is unnecessary, as it can be removed by redefining the notion of total energy. When the total energy includes the potential energy $-\rho g x$ due to gravity, the conservation form of the energy equation is restored. The source term in the momentum equation remains.

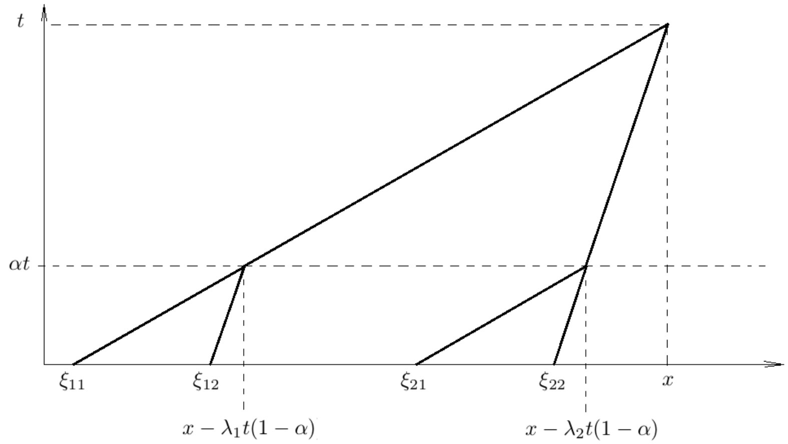

FIG. 2. *Illustration of the intermediate solutions and the involved characteristics for the first step in the Runge-Kutta scheme.*

fails to be accurate enough (the error is $\mathcal{O}(t^2)$ instead of $\mathcal{O}(t^3)$).

Recall the second order Runge-Kutta method for the ordinary differential equation

$$\dot{q}(t) = s(t, q(t)) \qquad q : \mathbb{R}_0^+ \to \mathbb{R} \tag{3.16}$$

$$q^{(1)}(\alpha t) = q(0) + \alpha t s(0, q(0)) \tag{3.17}$$

$$q(t) = q(0) + t\left(1 - \frac{1}{2\alpha}\right) s(0, q(0)) + t\frac{1}{2\alpha} s(\alpha t, q^{(1)}(\alpha t)) + \mathcal{O}(t^3) \tag{3.18}$$

for any $\alpha \in (0, 1)$. In particular choosing $\alpha = \frac{1}{2}$ (midpoint method) involves a predictor value at half time step. This can be taken as inspiration for constructing a sufficiently accurate approximate evolution operator:

THEOREM 3.2 (RK2 evolution operator). *Choose (see Figure 2)*

$$\xi_{\ell k} := x - \lambda_\ell t(1 - \alpha) - \lambda_k \alpha t \tag{3.19}$$

$$Q^*_{k\ell} := Q_{k,0}(\xi_{\ell k}) + \alpha t S_k(Q_{1,0}(\xi_{\ell k}), \dots, Q_{m,0}(\xi_{\ell k})) \qquad k, \ell = 1, \dots, m \tag{3.20}$$

and

(3.21)
$$Q^{(1)}_\ell(t, x) := Q_{\ell,0}(x - \lambda_\ell t) + \left(1 - \frac{1}{2\alpha}\right) S_\ell(Q_{1,0}(x - \lambda_\ell t), \dots, Q_{m,0}(x - \lambda_\ell t))t$$

$$+ \frac{t}{2\alpha} S_\ell\Big(Q^*_{1\ell}, \dots, Q^*_{m\ell}\Big) \qquad \ell = 1, \dots, m \tag{3.22}$$

Then, for all $\alpha \in (0, 1)$

$$Q^{(1)}_\ell(t, x) = Q_\ell(t, x) + \mathcal{O}(t^3) \qquad \ell = 1, \dots, m \tag{3.23}$$

Note that $Q^*_{\ell j}$ approximates $Q_\ell(\alpha t, x - \lambda_j t(1 - \alpha))$.

Proof. By explicitly computing the first three terms of the Taylor series in t one confirms the statement. The exact solution is

$$Q_\ell(t,x) = Q_{\ell,0}(x) + t\partial_t Q_\ell\Big|_{t=0} + \frac{t^2}{2}\partial_t^2 Q_\ell\Big|_{t=0} + \mathcal{O}(t^3) \tag{3.24}$$

$$= Q_{\ell,0}(x) + t(S_{\ell,0} - \lambda_\ell \partial_x Q_{\ell,0}) \tag{3.25}$$

$$+ \frac{t^2}{2}\left(\sum_k \frac{\partial S_\ell}{\partial Q_k}\Big(S_{k,0} - (\lambda_k + \lambda_\ell)\partial_x Q_{k,0}\Big) + \lambda_\ell^2 \partial_x^2 Q_{\ell,0})\right) + \mathcal{O}(t^3) \tag{3.26}$$

where $S_{\ell,0}$ denotes

$$S_{\ell,0} := S_\ell(Q_{1,0}(x), \ldots, Q_{m,0}(x)) \tag{3.27}$$

and $\frac{\partial S_\ell}{\partial Q_k}$ also is evaluated at x. Note that it has been used that $\partial_x \lambda_\ell = 0$ (i.e. that the homogeneous system is linear), but the source S can be any differentiable function of Q.

Expand now (3.22) ($\ell = 1, \ldots, m$):

$$\partial_t Q^*_{k\ell}\Big|_{t=0} = -(\lambda_\ell(1-\alpha) + \lambda_k \alpha)\partial_x Q_{k,0} + \alpha S_{k,0} \tag{3.28}$$

$$\partial_t Q_\ell^{(1)}(t,x) = -\lambda_\ell \partial_x Q_{\ell,0}(x - \lambda_\ell t) \tag{3.29}$$

$$+ \left(1 - \frac{1}{2\alpha}\right)\left(t\sum_k \frac{\partial S_\ell}{\partial Q_k}\partial_x Q_{k,0}(x - \lambda_\ell t)(-\lambda_\ell)\right. \tag{3.30}$$

$$\left. + S_\ell(Q_{1,0}(x - \lambda_\ell t), \ldots, Q_{m,0}(x - \lambda_\ell t))\right) \tag{3.31}$$

$$+ \frac{1}{2\alpha}\left(t\sum_k \frac{\partial S_\ell}{\partial Q_k}\partial_t Q^*_{k\ell} + S_\ell\Big(Q^*_{1\ell}, \ldots, Q^*_{m\ell}\Big)\right) \tag{3.32}$$

$$\overset{t=0}{=} -\lambda_\ell \partial_x Q_{\ell,0} + S_{\ell,0} \tag{3.33}$$

$$\partial_t^2 Q_\ell^{(1)}(t,x)\Big|_{t=0} = \lambda_\ell^2 \partial_x^2 Q_{\ell,0} + \left(1 - \frac{1}{2\alpha}\right)\left(2\sum_k \frac{\partial S_\ell}{\partial Q_k}\partial_x Q_{k,0}(-\lambda_\ell)\right) \tag{3.34}$$

$$+ \frac{1}{2\alpha}\left(2\sum_k \frac{\partial S_\ell}{\partial Q_k}\partial_t Q^*_{k\ell}\Big|_{t=0}\right) \tag{3.35}$$

$$= \lambda_\ell^2 \partial_x^2 Q_{\ell,0} - \sum_k \frac{\partial S_\ell}{\partial Q_k}\Big(\partial_x Q_{k,0}\,(\lambda_\ell + \lambda_k) - S_{k,0}\Big) \tag{3.36}$$

$\square$

Obviously the two Taylor series agree up to terms $\mathcal{O}(t^3)$, which proves the statement.

COROLLARY 3.3 (Midpoint method). *If $\alpha = \frac{1}{2}$, then for $\ell, k = 1, \ldots m$*

$$\xi_{\ell,j} := x - (\lambda_\ell + \lambda_j)\frac{t}{2} \tag{3.37}$$

$$Q^*_{k\ell} := Q_{k,0}(\xi_{\ell k}) + \frac{t}{2}S_k(Q_{1,0}(\xi_{k\ell}), \ldots, Q_{m,0}(\xi_{k\ell})) \tag{3.38}$$

$$Q_\ell^{(1)}(t,x) := Q_{\ell,0}(x - \lambda_\ell t) + tS_\ell\Big(Q^*_{1\ell}, \ldots, Q^*_{m\ell}\Big) \tag{3.39}$$

COROLLARY 3.4 (RK2 evolution operator for a scalar equation). *For a scalar equation*

$$(\partial_t + \lambda\partial_x)Q = S(Q) \tag{3.40}$$

the algorithm reads

$$\xi := x - \lambda t \tag{3.41}$$

and

$$Q^{(1)}(t,x) := Q_0(x-\lambda t) + \left(1 - \frac{1}{2\alpha}\right) S(Q_0(x-\lambda t))t \tag{3.42}$$

$$+ \frac{t}{2\alpha} S\Big(Q_0(\xi) + \alpha t S(Q_0(\xi))\Big) \tag{3.43}$$

For the equations (3.6)–(3.8) of linear acoustics with gravity, $\lambda_1 = c = -\lambda_2, \lambda_3 = 0$. The characteristic variables are

$$Q_1 = \frac{p+cv}{2} \qquad Q_2 = \frac{p-cv}{2} \qquad Q_3 = -\frac{p}{c^2} + \rho \tag{3.44}$$

and the gravity source term then is

$$S_1 = -S_2 = \frac{g}{2c}(Q_1+Q_2) + \frac{cg}{2}Q_3 \qquad S_3 = 0 \tag{3.45}$$

4. Update of the cell average in presence of a source term.

The update of the cell average needs to include the space-time average of the source term according to (2.3) of section 2.2. This space-time average needs to be approximated by a suitable quadrature / approximation with sufficient order of accuracy. Active flux has a strong focus on providing discrete degrees of freedom along the boundary which allow to perform a quadrature along the boundary. However, the evaluation of the source term for the update of the cell average involves an averaging over the cell volume. It is more difficult to achieve the desired order of accuracy here, as the setup lacks the quadrature points that would have been natural for this task. A quadrature formula adapted to the geometry of the active flux method is derived here.

Active flux for equations with a source term is considered in [NR16] for stationary problems, and for parabolic problems with slowly varying boundary conditions. In these cases there is no need to use high order quadrature in time. Therefore the method suggested there cannot be used here.

4.1. One spatial dimension.

The numerical discretization (2.5)

$$\hat{s}_{\mathcal{C}} \simeq \frac{1}{\Delta t}\int_{t^n}^{t^n+\Delta t} \mathrm{d}t\, \frac{1}{|\mathcal{C}|}\int_{\mathcal{C}} \mathrm{d}\mathbf{x}\, s(q(t,\mathbf{x})) \tag{4.1}$$

of the source term in (2.3) requires a space-time quadrature that is exact for parabolic functions. The natural candidate would be Simpson's rule in both space and time (as used for the numerical flux), but there are not enough quadrature points for it. For example in one spatial dimension, the available information is

$$
\begin{array}{c|ccc}
t^{n+1} & q^{n+1}_{i-\frac12} & & q^{n+1}_{i+\frac12} \\
t^{n+\frac12} & q^{n+\frac12}_{i-\frac12} & & q^{n+\frac12}_{i+\frac12} \\
t^{n} & q^{n+1}_{i-\frac12} & \boxed{\bar q^n_i} & q^{n}_{i+\frac12} \\
\hline
 & x_{i-\frac12} & & x_{i+\frac12}
\end{array}
$$

These are only 7 values (the box emphasizes that one of the values is a cell average, whereas the others are point values).

4.1.1. Linear source term. Consider first a linear source term, i.e. $s'' = 0$. Such source terms are relevant in practice (e.g. compressible Euler equations with gravity), and therefore it is worth dealing with them specifically as they allow for a simpler approach. For linear source it is possible to first find a quadrature for q and to apply s to the result. In order to find a quadrature formula for q, one needs to find a space-time polynomial $p(t, x)$ of at least second degree which interpolates the available 7 data. Integrating this polynomial would yield a quadrature formula for q. Here we suggest to use

$$
\mathscr{P}(t, x) = (a_0 + a_1 x + a_2 t + a_3 x^2 + a_4 xt + a_5 t^2) + a_6 x t^2 \tag{4.2}
$$

There is a unique set of coefficients $a_0, \ldots, a_6$ which makes polynomial (4.2) fulfill

$$
\mathscr{P}(t^{n+1}, x_{i-\frac12}) = q^{n+1}_{i-\frac12} \qquad \mathscr{P}(t^{n+1}, x_{i+\frac12}) = q^{n+1}_{i+\frac12} \tag{4.3}
$$

$$
\mathscr{P}(t^{n+\frac12}, x_{i-\frac12}) = q^{n+\frac12}_{i-\frac12} \qquad \mathscr{P}(t^{n+\frac12}, x_{i+\frac12}) = q^{n+\frac12}_{i+\frac12} \tag{4.4}
$$

$$
\mathscr{P}(t^{n}, x_{i-\frac12}) = q^{n}_{i-\frac12} \qquad \int\limits_{x_{i-\frac12}}^{x_{i+\frac12}} \mathrm{d}x\, \mathscr{P}(t^n, x) = q^n_i \qquad \mathscr{P}(t^{n}, x_{i+\frac12}) = q^{n}_{i+\frac12} \tag{4.5}
$$

Inserting this polynomial in (2.5) and integrating it instead of the source yields the following quadrature formula:

$$
\begin{aligned}
&\frac{1}{\Delta t}\int_0^{\Delta t} \mathrm{d}t\, \frac{1}{\Delta x}\int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \mathrm{d}x\, q(t^n + t, x_i + x) = \\
&\bar q^n_i + \frac{1}{12}\left(-5(q^n_{i-\frac12} + q^n_{i+\frac12}) + q^{n+1}_{i-\frac12} + q^{n+1}_{i+\frac12} + 4(q^{n+\frac12}_{i-\frac12} + q^{n+\frac12}_{i+\frac12})\right)
\end{aligned} \tag{4.6}
$$

The weights can be depicted as

$$
\begin{array}{c|ccc}
t^{n+1} & \frac{1}{12} & & \frac{1}{12} \\
t^{n+\frac12} & \frac{4}{12} & & \frac{4}{12} \\
t^{n} & -\frac{5}{12} & \boxed{1} & -\frac{5}{12} \\
\hline
 & x_{i-\frac12} & & x_{i+\frac12}
\end{array}
$$

Again, the box indicates that the corresponding weight refers to the cell average, whereas the others multiply point values.

The time levels $(n, n+\frac12, n+1)$ contribute with weights $\left(\frac16, \frac23, \frac16\right)$, such that this quadrature formula is a modification of Simpson's rule in time. Note that it is not possible to use terms proportional to x^3, x^2t or t^3 instead of the term xt^2 in the polynomial ansatz, as then the system (4.3)–(4.5) does not admit a solution. In a sense this is therefore the only choice of a simple quadrature formula.

Quadrature formula (4.6) can be used immediately in order to approximate (2.5) for linear source terms.

4.1.2. Nonlinear source term. For nonlinear s, the average

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathrm{d}x\, s(q(t^n, x)) \tag{4.7}$$

in general is different from

$$s\left(\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathrm{d}x\, q(t^n, x)\right) \tag{4.8}$$

Point values, however, do not present any difficulties: one can just evaluate s on them. Therefore we suggest to consider a reconstruction $q_{\mathrm{recon},i}(x)$ that interpolates $q^n_{i-\frac{1}{2}}$ and $q^n_{i+\frac{1}{2}}$ and whose average agrees with $\bar{q}^n_i$. It is computed anyway in order to update the point values in time, see equation (2.7). This reconstruction can be easily evaluated at the midpoint of the cell. Then, instead of the cell averages, one works with a seventh point value $q_{\mathrm{recon},i}(0) = \frac{1}{4}(6\bar{q}^n_i - q^n_{i-\frac{1}{2}} - q^n_{i+\frac{1}{2}})$. Of course, this is equivalent to replacing the average by a Simpson's rule in the quadrature, and thus the order of the quadrature is not reduced. Therefore when using only point values (the 6 pointwise degrees of freedom and one value at the cell midpoint) the weights of the quadrature formula read

$$\begin{array}{c|ccc} t^{n+1} & \frac{1}{12} & & \frac{1}{12} \\ t^{n+\frac{1}{2}} & \frac{4}{12} & & \frac{4}{12} \\ t^n & -\frac{3}{12} & \frac{8}{12} & -\frac{3}{12} \\ \hline & x_{i-\frac{1}{2}} & & x_{i+\frac{1}{2}} \end{array}$$

Equation (2.5) then is replaced by the quadrature

$$\hat{s}_i = \frac{s(q^{n+1}_{i-\frac{1}{2}}) + s(q^{n+1}_{i+\frac{1}{2}}) + 4\Big(s(q^{n+1}_{i-\frac{1}{2}}) + s(q^{n+1}_{i+\frac{1}{2}})\Big) - 3\Big(s(q^{n+1}_{i-\frac{1}{2}}) + s(q^{n+1}_{i+\frac{1}{2}})\Big) + 8q_{\mathrm{recon},i}(0)}{12} \tag{4.9}$$

This quadrature can now be used for nonlinear s. As (4.9) uses a Simpson quadrature instead of the average, upon usage of a linear source s, it reduces to the expression (4.6) because of the quadratic reconstruction.

If the source term vanishes, the scheme becomes conservative in the sense that averages are updated using numerical fluxes.

4.2. Two spatial dimensions.

4.2.1. Linear source term. Similarly consider the setup of the active flux method on two-dimensional Cartesian grids as described in 2.1. The available de-

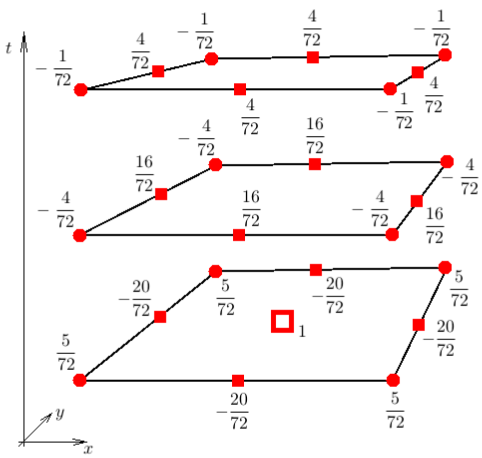

FIG. 3. *Illustration of the weights of the space time quadrature formula* (4.15).

grees of freedom are

$$3 \times 4 \text{ nodes: } q^n_{i\pm\frac{1}{2},j\pm\frac{1}{2}}, q^{n+\frac{1}{2}}_{i\pm\frac{1}{2},j\pm\frac{1}{2}}, q^{n+1}_{i\pm\frac{1}{2},j\pm\frac{1}{2}} \tag{4.10}$$

$$3 \times 2 \text{ vertical edges: } q^n_{i\pm\frac{1}{2},j}, q^{n+\frac{1}{2}}_{i\pm\frac{1}{2},j}, q^{n+1}_{i\pm\frac{1}{2},j} \tag{4.11}$$

$$3 \times 2 \text{ horizontal edges: } q^n_{i,j\pm\frac{1}{2}}, q^{n+\frac{1}{2}}_{i,j\pm\frac{1}{2}}, q^{n+1}_{i,j\pm\frac{1}{2}} \tag{4.12}$$

$$1 \text{ average: } \bar{q}^n_{ij} \tag{4.13}$$

The ansatz for a space-time polynomial is

$$\mathscr{P}(t,x,y) = \left(\sum_{\zeta+\eta+\vartheta \le 4} a_{\zeta\eta\vartheta} \cdot x^\zeta y^\eta t^\vartheta \right) + a_{212} x^2 y t^2 + a_{122} x y^2 t^2 \tag{4.14}$$

It admits a unique solution to the interpolation problem given the available degrees of freedom and yields the following quadrature formula (see also figure 3):

$$\begin{aligned}
&\frac{1}{\Delta x} \int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \mathrm{d}x \, \frac{1}{\Delta y} \int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \mathrm{d}y \, \frac{1}{\Delta t} \int_0^{\Delta t} \mathrm{d}t \, q(t,x,y) = \bar{q}^n_{ij} \\
&- \frac{20}{72} \left(q^n_{\mathrm{E}} + q^n_{\mathrm{N}} + q^n_{\mathrm{S}} + q^n_{\mathrm{W}} \right) + \frac{5}{72} \left(q^n_{\mathrm{N}E} + q^n_{\mathrm{N}W} + q^n_{\mathrm{S}E} + q^n_{\mathrm{S}W} \right) \\
&+ \frac{16}{72} \left(q^{n+\frac{1}{2}}_{\mathrm{E}} + q^{n+\frac{1}{2}}_{\mathrm{N}} + q^{n+\frac{1}{2}}_{\mathrm{S}} + q^{n+\frac{1}{2}}_{\mathrm{W}} \right) - \frac{4}{72} \left(q^{n+\frac{1}{2}}_{\mathrm{N}E} + q^{n+\frac{1}{2}}_{\mathrm{N}W} + q^{n+\frac{1}{2}}_{\mathrm{S}E} + q^{n+\frac{1}{2}}_{\mathrm{S}W} \right) \\
&+ \frac{4}{72} \left(q^{n+1}_{\mathrm{E}} + q^{n+1}_{\mathrm{N}} + q^{n+1}_{\mathrm{S}} + q^{n+1}_{\mathrm{W}} \right) - \frac{1}{72} \left(q^{n+1}_{\mathrm{N}E} + q^{n+1}_{\mathrm{N}W} + q^{n+1}_{\mathrm{SE}} + q^{n+1}_{\mathrm{SW}} \right)
\end{aligned} \tag{4.15}$$

The time levels $(n, n+\frac{1}{2}, n+1)$ contribute again with weights $(\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$, and the edges always contribute -4 times the nodes.

4.2.2. Nonlinear source term. Again, for nonlinear source instead of the average it is necessary to use the evaluation of the reconstruction at the cell midpoint.

This amounts to an approximation of the average by a two-dimensional Simpson rule. Then the source term is approximating as follows:

$$\begin{aligned}
\frac{1}{\Delta x}\int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \mathrm{d}x\, \frac{1}{\Delta y}\int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \mathrm{d}y\, \frac{1}{\Delta t}\int_0^{\Delta t} \mathrm{d}t\, s(q(t,x,y)) = &\frac{32}{72} s(q_{\mathrm{recon},ij}(0,0))\\
&- \frac{12}{72}\left(s(q^n_{\mathrm{E}}) + s(q^n_{\mathrm{N}}) + s(q^n_{\mathrm{S}}) + s(q^n_{\mathrm{W}})\right)\\
&+ \frac{3}{72}\left(s(q^n_{\mathrm{N}E}) + s(q^n_{\mathrm{N}W}) + s(q^n_{\mathrm{S}E}) + s(q^n_{\mathrm{S}W})\right)\\
&+ \frac{16}{72}\left(s(q^{n+\frac12}_{\mathrm{E}}) + s(q^{n+\frac12}_{\mathrm{N}}) + s(q^{n+\frac12}_{\mathrm{S}}) + s(q^{n+\frac12}_{\mathrm{W}})\right)\\
&- \frac{4}{72}\left(s(q^{n+\frac12}_{\mathrm{N}E}) + s(q^{n+\frac12}_{\mathrm{N}W}) + s(q^{n+\frac12}_{\mathrm{S}E}) + s(q^{n+\frac12}_{\mathrm{S}W})\right)\\
&+ \frac{4}{72}\left(s(q^{n+1}_{\mathrm{E}}) + s(q^{n+1}_{\mathrm{N}}) + s(q^{n+1}_{\mathrm{S}}) + s(q^{n+1}_{\mathrm{W}})\right)\\
&- \frac{1}{72}\left(s(q^{n+1}_{\mathrm{N}E}) + s(q^{n+1}_{\mathrm{N}W}) + s(q^{n+1}_{\mathrm{S}E}) + s(q^{n+1}_{\mathrm{S}W})\right)
\end{aligned} \tag{4.16}$$

In case that the data only depend on one of the variables, the two-dimensional quadratures (4.15) and (4.16) do *not* exactly reduce to the one dimensional quadratures (4.6) and (4.9). This is because (cf. Figure 3) the point values on edge midpoints $\left(0, \pm\frac{\Delta y}{2}\right)$ do not disappear even if the data depend only on x, and therefore the available degrees of freedom remain different from the one-dimensional case.

5. Well-balanced property for acoustics with gravity.

5.1. Exact evolution operator. As described in 3.2 a closed-form exact evolution operator for acoustics with gravity is very difficult to obtain. Nevertheless, it is still possible to show that a scheme endowed with such an operator would be well-balanced / stationarity preserving; i.e. that there exists a discretization of the stationary states of the PDE which remain exactly stationary. This proof does not require the evolution operator to be known explicitly, but only relies on the fact that it is exact. Besides its fundamental importance, this result is used in section 5.2 to analyze the situation for the approximate evolution operator and to restore the well-balanced property for it.

The numerical stationary states are best studied upon the (discrete) Fourier transform. Define $t_x := \exp(\mathbb{i} k_x \Delta x)$, $t_y := \exp(\mathbb{i} k_y \Delta y)$. Here $\mathbb{i}$ is the imaginary unit and $\mathbf{k} = (k_x, k_y) \in \mathbb{R}^2$ is the wave vector characterizing the spatial frequency of the Fourier mode. Applying the Fourier transform introduces one mode $\bar q$ for the averages and one mode q for the point values; this implies writing $q_i := \bar q t_x^i t_y^j$, $q_{i+\frac12} := q t_x^i t_y^j$.

THEOREM 5.1 (Stationarity preservation with exact evolution). *If the discrete data fulfill*

$$\bar\rho_i = \frac{\rho_{i+\frac12} + \rho_{i-\frac12}}{2} \tag{5.1}$$

$$\frac{p_{i+\frac12} - p_{i-\frac12}}{\Delta x} = g\frac{\rho_{i-\frac12} + \rho_{i+\frac12}}{2} \tag{5.2}$$

$$\frac{\bar p_{i+\frac32} - \bar p_{i+\frac12}}{\Delta x} = g\frac{\rho_{i+\frac32} + 4\rho_{i+\frac12} + \rho_{i-\frac12}}{6} \tag{5.3}$$

and the exact evolution operator for (3.6)–(3.8) *is used, then the numerical solution remains stationary.*

Proof. The proof consists of two parts.

i) Consider first the evolution of the point values. When the exact evolution operator is used to update the point values, they remain stationary if the reconstruction fulfills

$$v_{\mathrm{recon}}(x) = \mathrm{const} \qquad\qquad \partial_x p_{\mathrm{recon}}(x) = \rho_{\mathrm{recon}}(x) g \tag{5.4}$$

Upon the Fourier transform this becomes (w.l.o.g. $x_i = 0$)

$$-3\left(2\bar{p} - p\left(1 + \frac{1}{t_x}\right)\right)\frac{2x}{\Delta x^2} + p\left(1 - \frac{1}{t_x}\right)\frac{1}{\Delta x} = \tag{5.5}$$

$$-3g\left(2\bar{\rho} - \rho\left(1 + \frac{1}{t_x}\right)\right)\frac{x^2}{\Delta x^2} + g\rho\left(1 - \frac{1}{t_x}\right)\frac{x}{\Delta x} + g\frac{6\bar{\rho} - \rho\left(1 + \frac{1}{t_x}\right)}{4} \tag{5.6}$$

This shall be valid for all x:

$$2\bar{\rho} - \rho(1 + 1/t_x) = 0 \tag{5.7}$$

$$-2\bar{p}t_x + p(t_x + 1) = \frac{\Delta x g \rho (t_x - 1)}{6} \tag{5.8}$$

$$p(t_x - 1) = \Delta x g \frac{6\bar{\rho} t_x - \rho(t_x + 1)}{4} \tag{5.9}$$

These are three equations for four variables. In particular

$$\bar{\rho} = \frac{\rho(1 + 1/t_x)}{2} \tag{5.10}$$

$$p = \Delta x g \rho \frac{t_x + 1}{2(t_x - 1)} \tag{5.11}$$

$$\bar{p} = \Delta x g \rho \frac{t_x^2 + 4t_x + 1}{6t_x(t_x - 1)} \tag{5.12}$$

These statements can be rewritten as finite difference formulae by inverting the Fourier transform:

$$\bar{\rho} = \frac{\rho_{i+\frac{1}{2}} + \rho_{i-\frac{1}{2}}}{2} \tag{5.13}$$

$$\frac{p_{i+\frac{1}{2}} - p_{i-\frac{1}{2}}}{\Delta x} = g\frac{\rho_{i-\frac{1}{2}} + \rho_{i+\frac{1}{2}}}{2} \tag{5.14}$$

$$\frac{\bar{p}_{i+1} - \bar{p}_i}{\Delta x} = g\frac{\rho_{i+\frac{3}{2}} + 4\rho_{i+\frac{1}{2}} + \rho_{i-\frac{1}{2}}}{6} \tag{5.15}$$

ii) Assume now (5.10)–(5.12) to be true. Simpson's rule in time for the flux average is trivial, and thus the update of the cell average amounts to

$$\frac{\bar{v}^{n+1} - \bar{v}^n}{\Delta t} + \frac{p(1 - 1/t_x)}{\Delta x} = \frac{\bar{v}^{n+1} - \bar{v}^n}{\Delta t} + g\rho\frac{t_x + 1}{2t_x} \tag{5.16}$$

$$= \frac{\bar{v}^{n+1} - \bar{v}^n}{\Delta t} + g\bar{\rho} \tag{5.17}$$

∎

The quadrature formula (4.6) for the source reduces to $g\bar{\rho}$ if the point values are stationary, which implies $\bar{v}^{n+1} = \bar{v}^n$. This completes the proof.

The equations (5.10)–(5.12) contain ρ as a free variable. One can rewrite the system making p the free variable:

$$\bar\rho = \frac{p(t_x-1)}{t_x\Delta x g} \qquad \rho = \frac{2p(t_x-1)}{\Delta x g(t_x+1)} \qquad \bar p = p\frac{t_x^2+4t_x+1}{3t_x(t_x+1)} \tag{5.18}$$

This form will be useful later.

Equations (5.2)–(5.3) are finite difference approximations of $\partial_x p = \rho g$. Equation (5.1) implies that the reconstructed ρ of the discrete stationary state is linear, which is clear: for quadratic reconstructions to fulfill (5.4), ρ_{recon} has to be linear in each cell. The slope of the linear function can vary from cell to cell and is given by (5.2).

5.2. Approximate evolution operator. The above section identifies conditions (5.1)–(5.3) on the discrete data for them to remain stationary upon usage of the *exact* evolution operator. Unfortunately, such an operator is unavailable in practice. Having identified an approximate solution operator, which agrees with the exact solution up to terms $\mathcal{O}(t^3)$ in section 3.3, here we study whether it keeps the same data (5.1)–(5.3) stationary as well.

THEOREM 5.2. *If the discrete data fulfill* (5.1)–(5.3) *and the approximate evolution operator of theorem 3.2 for* (3.6)–(3.8) *is used, then both the pressure p and the density ρ remain stationary over one time step, but the velocity undergoes the time evolution*

$$v_{i+\frac12}(t) = -\frac{\alpha g^2}{4}\frac{\rho_{i+\frac12}-\rho_{i-\frac12}}{\Delta x}t^3 \tag{5.19}$$

Proof. Assume the initial data to fulfill (5.1)–(5.3), or equivalently (5.4). Using (2.7) (and applying the discrete Fourier transform straight away) (5.4) implies

$$p_{\text{recon}}(x) = \frac14\left(6\bar p - p\left(1+\frac{1}{t_x}\right)\right) + \frac{x}{\Delta x}\left(1-\frac{1}{t_x}\right)p - 3\frac{x^2}{\Delta x^2}\left(2\bar p - p\left(1+\frac{1}{t_x}\right)\right) \tag{5.20}$$

$$\rho_{\text{recon}}(x) = \frac{1}{g\Delta x}\left(p\left(1-\frac{1}{t_x}\right) - 6\frac{x}{\Delta x}\left(2\bar p - p\left(1+\frac{1}{t_x}\right)\right)\right) \tag{5.21}$$

$$v_{\text{recon}}(x) = 0 \tag{5.22}$$

and using (3.44) therefore

$$Q_{1,0}(x) = Q_{2,0}(x) = -\frac{p(1+t_x)-6\bar p t_x}{8t_x} + \frac{p(t_x-1)x}{2\Delta x t_x} + \frac{3(p(1+t_x)-2\bar p t_x)x^2}{2\Delta x^2 t_x} \tag{5.23}$$

$$\begin{aligned} Q_{3,0}(x) = {} & \frac{p(-1+t_x)}{\Delta x g t_x} + \frac{p - 6\bar p t_x + p t_x}{4c^2 t_x} \\ & + \frac{\left(-\Delta x g p(t_x-1) + 6c^2(p(1+t_x)-2\bar p t_x)\right)x}{c^2\Delta x^2 g t_x} - \frac{3(p(1+t_x)-2\bar p t_x)x^2}{c^2\Delta x^2 t_x} \end{aligned} \tag{5.24}$$

Evaluating the Runge-Kutta algorithm of section 3.3 on these initial data (at

$x = \frac{\Delta x}{2}$) yields

$$\begin{pmatrix} \rho \\ v^* \\ p \end{pmatrix} \qquad \text{with} \qquad v^* = -\frac{\alpha g (t_x - 1)^2}{2\Delta x^2 t_x (t_x + 1)} p t^3 \tag{5.25}$$

(α is the parameter appearing in the RK2 method.)

Recall that ρ and p are the Fourier coefficients of the point values of the density and the pressure. Obviously ρ and p remain stationary, but the velocity does not. Using (5.18) v^* can be rewritten as

$$v^* = -\frac{\alpha g^2}{4\Delta x}\left(1 - \frac{1}{t_x}\right)\rho t^3 = -\frac{\alpha g^2}{4}\frac{\rho_{i+\frac{1}{2}} - \rho_{i-\frac{1}{2}}}{\Delta x} t^3 \tag{5.26}$$

having applied the inverse Fourier transform in the last step. ∎

Observe that the time evolution of the velocity is consistent with the accuracy of the algorithm ($\mathcal{O}(t^3)$).

COROLLARY 5.3 (Stationarity preservation with approximate evolution). *If the algorithm of section 3.3 is modified by adding the term*

$$\frac{\alpha g^2}{4}\frac{\rho_{i+\frac{1}{2}} - \rho_{i-\frac{1}{2}}}{\Delta x} t^3 \tag{5.27}$$

to the velocity evolution, then
i) its accuracy is not changed
ii) it becomes stationarity preserving / well-balanced with the same discrete stationary states as the exact evolution operator

The two forms (5.25) and (5.19) of v^* are equivalent, because the initial data have been chosen to be stationary, and thus additionally fulfill (5.18). The proposed modification is to *always* add $-v^*$ to the velocity evolution, irrespective of whether the data fulfill (5.18) or not. At this point the Fourier coefficients of ρ and p are independent and it matters whether the correction is used in the form (5.25) or (5.19). Of course, also the inverse Fourier transform has to be applied to the expression first in order for the correction to attain the form of a finite difference formula. Compact finite difference formulae are in one-to-one-correspondence with Laurent polynomials in t_x. An expression such as $\frac{1}{t_x+1} = 1 - t_x + t_x^2 \mp \dots$ is an expression involving an unbounded stencil and cannot be implemented in usual codes. Therefore (5.19) cannot be used as a correction because the correction would have a non-compact stencil (just as the equivalent expressions involving only $\bar{\rho}$ or $\bar{p}$). This is why the form (5.25) which involves point values of ρ is preferred.

Being always present in the velocity evolution (and not only at stationary states), the modification (5.27) might in general affect the stability of the algorithm, but it has not been found to have any effect on the stability in practice.

6. Numerical examples. The numerical examples of this section serve to illustrate the performance of the new method. The equations discussed are linear advection with different source terms (in one and two spatial dimensions, as introduced in section 3.1) and linear acoustics with gravity (introduced in section 3.2). In both cases it is demonstrated that the method achieves third order of accuracy in the experiments. For acoustics with gravity additionally the discrete stationary states are studied and shown to agree with the prediction of section 5.

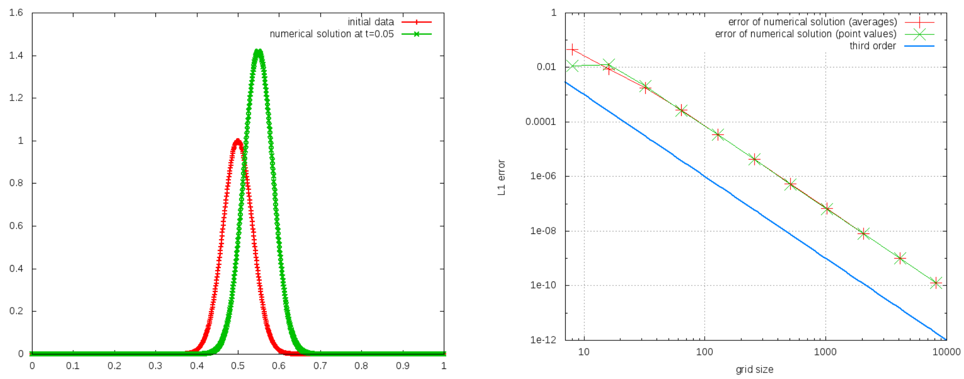

FIG. 4. *Gaussian initial data for* (6.1) *with* $\mathbf{U} = \mathbf{e}_x$*,* $\kappa = 7$*. Note that due to the source term, the Gaussian is advected and also changes shape. Exact evolution operator* (3.4) *and quadrature formula* (4.6) *have been used with CFL = 0.45.* Left*: Initial data and solution at* $t = 0.05$ *(cell averages) on a grid with 1000 cells.* Right*: Error of the numerical solution as a function of the grid size shows third order convergence.*

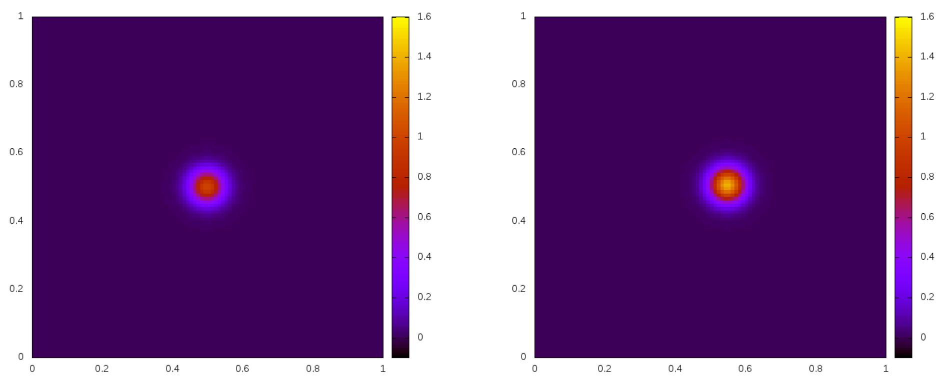

FIG. 5. *Gaussian initial data for* (6.1) *with* $\mathbf{U} = (1, 0.1)$*,* $\kappa = 7$*. Note that due to the source term, the Gaussian is advected and also changes shape. Exact evolution operator* (3.4) *and quadrature formula* (4.15) *have been used with CFL = 0.45.* Left*: Initial setup.* Right*: Numerical solution at* $t = 0.05$ *on a* 100×100 *Cartesian grid.*

6.1. Linear advection. Consider first

$$\partial_t q + \mathbf{U} \cdot \nabla q = \kappa q \tag{6.1}$$

with the exact solution given by (3.4). In Figures 4–6 the exact solution operator is used for the evolution of the point values and third order convergence is observed. This shows that the quadrature formulae (4.6) and (4.15) used to evolve the cell averages indeed yield a third order scheme. Figure 4 shows the setup for a one-dimensional situation together with a convergence study, Figure 5 shows the setup in two spatial dimensions and Figure 6 shows the corresponding convergence study.

Consider now

$$\partial_t q + \mathbf{U} \cdot \nabla q = \kappa q^B \qquad B \neq 1 \tag{6.2}$$

with the exact solution (3.5) and $\kappa = 7$, $B = 3$. Figure 7 (left) shows the initial data and the numerical solution, and Figure 7 (right) shows a convergence study for

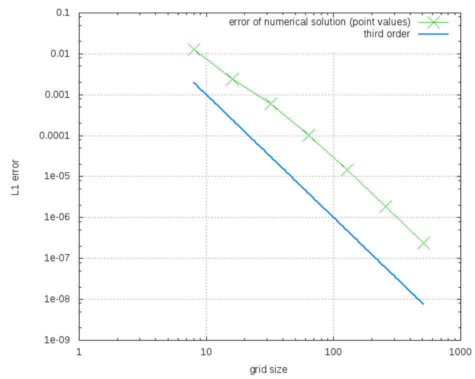

FIG. 6. *Convergence study for the setup shown in Figure 5. One observes third order accuracy.*

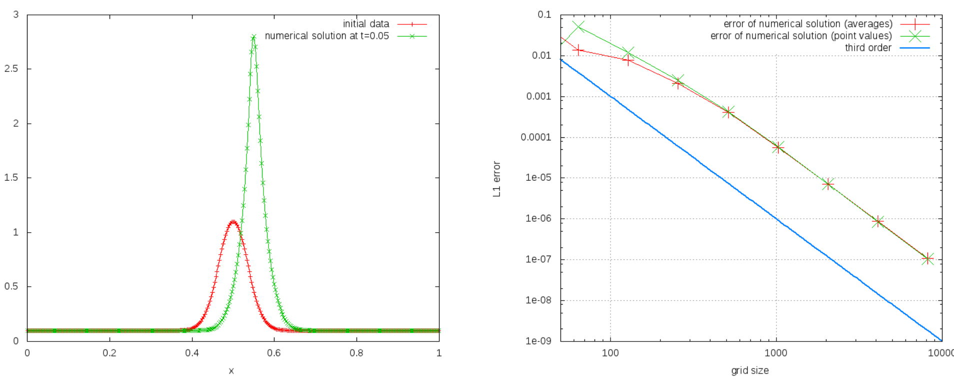

FIG. 7. *Gaussian initial data for* (6.2) *with* $s(q) = \kappa q^B$ *and* $\mathbf{U} = \mathbf{e}_x$, $\kappa = 7$, $B = 3$. *Runge-Kutta approximate evolution operator from Corollary 3.4 and quadrature formula* (4.9) *have been used with CFL = 0.45. The solution has been computed on a grid covering* $[-1 : 2]$, *but the error is only computed inside* $[0, 1]$ *to exclude any boundary influence.* Left: *Initial setup and solution at* $t = 0.05$ *(point values) on a grid with 1000 cells.* Right: *Error of the numerical solution as a function of the grid size shows third order convergence. The exact solution is given by* (3.5).

the approximate evolution operator from Corollary (3.4). One observes third order accuracy, as expected.

6.2. Acoustics with gravity. Consider now the equations of linear acoustics with a gravity source term (3.6)–(3.8). The exact solution operator is only partly available in closed form, and therefore the approximate Runge-Kutta evolution operator of section 3.3 is used in combination with the well-balancing fix (5.27). The parameter α in the Runge-Kutta method is chosen to $\alpha = \frac{1}{2}$.

Figure 8 shows a stationary setup given by

$$p = A_1 x^2 + A_2 x + A_3 \qquad \rho = 2A_1 x/g + A_2/g \qquad v = 0 \tag{6.3}$$

with $A_1 = 17, A_2 = -3, A_3 = 1$. This parabola is exactly recovered by the reconstruction, and thus remains stationary up to machine precision. This experiment shows that the well-balancing fix works as it should.

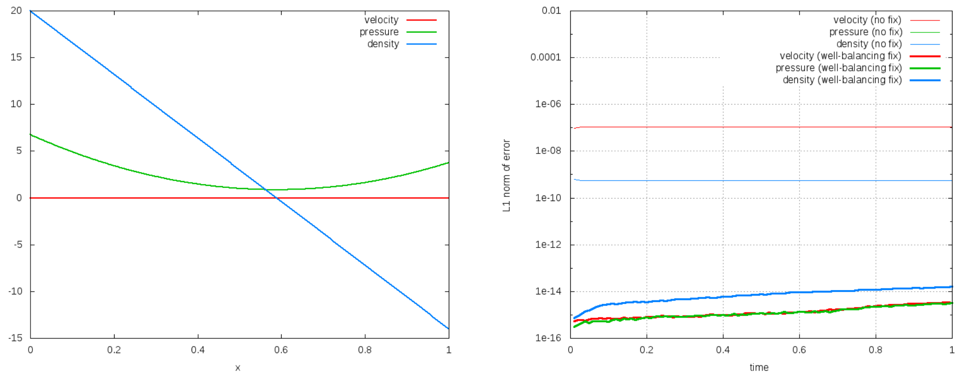

FIG. 8. *Setup of a stationary parabola* (6.3) *for* (3.6)–(3.8)*, solved using the Runge-Kutta approximate evolution operator of section 3.3 with and without well-balancing* (5.27)*. Here* $g = -1$*, and the setup is solved on a grid covering* $[-1.5, 2.5]$*, but the error is only measured inside* $[0, 1]$ *(*$\Delta x = 10^{-2}$*) to exclude the influence of the boundaries.* Left*: Setup.* Right*: Error of numerical solution as a function of time. Thin lines: without the well-balancing* (5.27)*. Thick lines: including the well-balancing* (5.27)*. In the latter case one only observes an evolution due to machine error.*

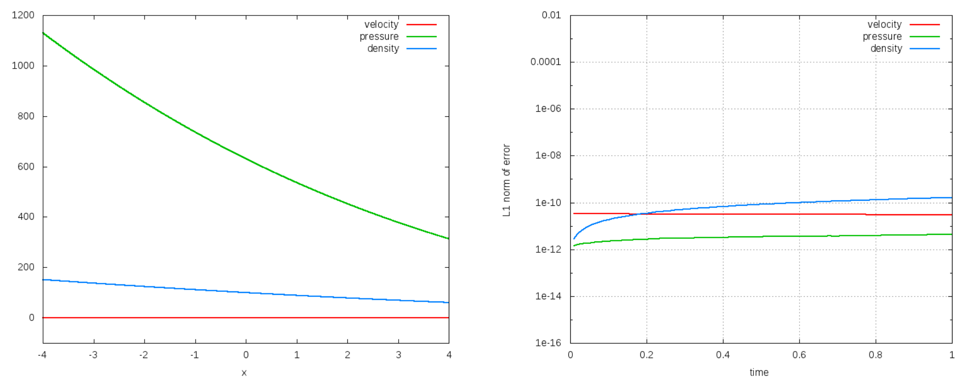

FIG. 9. *Stationary setup* (6.4) *for* (3.6)–(3.8)*, solved using the Runge-Kutta approximate evolution operator of section 3.3 with well-balancing* (5.27)*. Here* $g = -1$*, and the setup is solved on a grid covering* $[-5.5, 5.5]$*, but the error is only measured inside* $[-4, 4]$ *(*$\Delta x = 10^{-3}$*) to exclude the influence of the boundaries.* Left*: Setup (cell averages).* Right*: Error of numerical solution as a function of time. One observes a transition towards a numerical stationary state which then persists forever.*

Consider next (Figure 9) the stationary setup fulfilling $p = K\rho^\gamma$, i.e.

$$\rho = \left(\frac{g(\gamma - 1)}{K\gamma} x + \rho_0^{\gamma-1} \right)^{\frac{1}{\gamma-1}} \tag{6.4}$$

with $K = 1, \gamma = 1.4$, $\rho_0 = 100$. This is reminiscent of an isentropic atmosphere in the context of the Euler equations. This setup is not recovered exactly by the reconstruction, but one observes a numerical evolution towards a discrete stationary state which then persists forever.

Next, a perturbation

$$200 \exp(-100x^2) \tag{6.5}$$

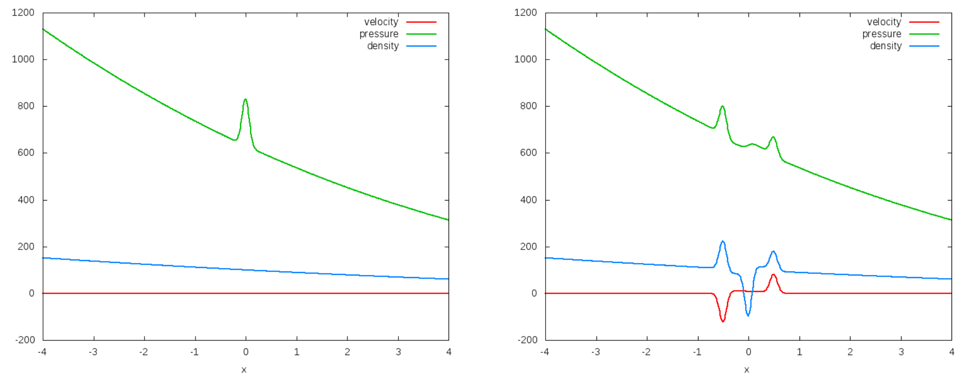

FIG. 10. *Setup* (6.4) *endowed with the pressure perturbation* (6.5) *solved using the Runge-Kutta approximate evolution operator of section 3.3 with well-balancing* (5.27)*.* Left*: Initial data (cell averages).* Right*: Numerical solution (cell averages) at* $t = 0.5$ *on a grid covering* $[-5.5, 5.5]$*, but only the subinterval* $[-4, 4]$ *is considered in order to exclude the influence of the boundaries.* $\Delta x = 0.01$*, CFL = 0.45.*

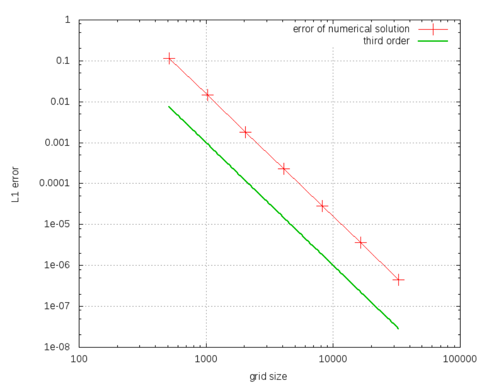

FIG. 11. *Setup of Figure 10. The error of the numerical solution is measured on the point values. One observes third order accuracy.*

in the pressure is added onto the setup (6.4). In order to study the accuracy of the scheme on this setup, it is solved on a grid of $131072 = 2^{18}$ cells and the solution is used as reference. Again, $g = -1, K = 1, \gamma = 1.4$. Figure 10 shows the setup and the numerical solution at $t = 0.5$, and Figure 11 shows a convergence study which displays third order convergence.

Consider finally a Riemann problem:

$$\rho = 3.5 \qquad p = 1.5 \qquad v = \begin{cases} 1 & 0.25 \le x \le 0.75 \\ 3 & \text{else} \end{cases} \tag{6.6}$$

This Riemann problem can be solved exactly using the formula (A.18)–(A.22). Note that if all quantities are constant in space, then they solve

$$\partial_t \rho = 0 \tag{6.7}$$
$$\partial_t p = 0 \tag{6.8}$$
$$\partial_t v = \rho g \tag{6.9}$$

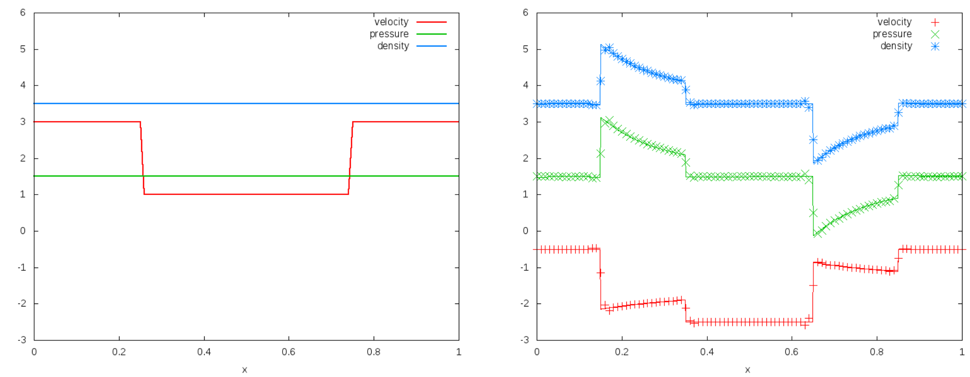

FIG. 12. *Riemann problem setup* (6.6) *solved using the Runge-Kutta approximate evolution operator of section 3.3 with well-balancing* (5.27). *Here,* $g = -10$. Left*: Initial data.* Right*: Numerical solution (dots) and exact solution (solid line) at* $t = 0.1$. $\Delta x = 0.01$*, CFL = 0.45. Point values of the numerical solution are shown are shown.*

which means that ρ and p remain stationary, but that $v = v(t = 0) + \rho g t$. The solution to the initial data (6.6) therefore can be obtained by adding the time evolution of $\begin{pmatrix} 0 \\ v_0(x) \\ 0 \end{pmatrix}$ (via numerical quadrature of (A.18)–(A.22)) and the time evolution of $\begin{pmatrix} \rho \\ 0 \\ p \end{pmatrix}$ which is just $\begin{pmatrix} \rho \\ \rho g t \\ p \end{pmatrix}$. Figure 12 shows the numerical and the exact solution.

7. Conclusions and outlook.

Active flux is a novel kind of numerical method for hyperbolic problems, extending the finite volume method. Instead of computing the intercell flux via a Riemann problem it relies on a continuous reconstruction and on accurately evolved point values along the cell boundary. They then immediately serve as quadrature values for the computation of the intercell flux. The extension of active flux to time dependent balance laws presented in this paper requires a modification in both these aspects: the evolution of the point values and the average update need to account for the source term. Here, an approximate evolution operator is suggested for the point value update; this is done for linear systems with possibly nonlinear source terms in one spatial dimension, and linear scalar equations with source terms in multiple spatial dimensions. A suitable quadrature is suggested in order to approximate the contribution of the source term to the cell average. This quadrature can be applied to any system of (nonlinear) balance laws.

We aim at combining the strategy presented in this paper with an approximate evolution operator for a nonlinear homogeneous problem (such as those suggested in [Bar19a]) in future. Multi-dimensional systems of hyperbolic conservation laws are very different from their one-dimensional counterparts because in general characteristics are unavailable and need to be conceptually replaced by characteristic cones. Examples of evolution operators that make use of such cones can be found in [ER13, FR15, Fan17, BHKR19]. Combining these with an approximate evolution of the source term shall pave the way towards the extension of active flux to nonlinear multi-dimensional balance laws and the derivation of accurate structure preserving (in particular well-balanced) methods for them.

Appendix A. Exact solution of linear acoustics with gravity.

System (3.6)–(3.8) can in principle be immediately solved exactly via Fourier transform by inserting the ansatz

$$\begin{pmatrix} \rho \\ v \\ p \end{pmatrix} = \begin{pmatrix} \hat\rho \\ \hat v \\ \hat p \end{pmatrix} \exp(\mathbb{i}k \cdot x - \mathbb{i}\omega t) \tag{A.1}$$

into (3.6)–(3.8):

$$\omega \begin{pmatrix} \hat\rho \\ \hat v \\ \hat p \end{pmatrix} = \begin{pmatrix} 0 & k & 0 \\ \mathbb{i}g & 0 & k \\ 0 & c^2k & 0 \end{pmatrix} \begin{pmatrix} \hat\rho \\ \hat v \\ \hat p \end{pmatrix} \tag{A.2}$$

Therefore $\omega = 0$, or $\omega = \pm\sqrt{c^2k^2 + \mathbb{i}gk}$. The complex eigenvalue can be removed upon transforming

$$\rho = \tilde\rho e^{\mu x} \qquad v = \tilde v e^{\mu x} \qquad p = \tilde p e^{\mu x} \tag{A.3}$$

with

$$\mu := \frac{g}{2c^2} \tag{A.4}$$

System (3.6)–(3.8) then reads

$$\partial_t \tilde\rho + \partial_x \tilde v = -\mu \tilde v \tag{A.5}$$

$$\partial_t \tilde v + \partial_x \tilde p = \tilde\rho g - \mu \tilde p \tag{A.6}$$

$$\partial_t \tilde p + c^2 \partial_x \tilde v = -c^2 \mu \tilde v \tag{A.7}$$

Now, a solution of (A.5)–(A.7) shall be found. For better readability, drop the tilde. Upon the Fourier transform (A.5)–(A.7) becomes

$$\omega \begin{pmatrix} \hat\rho \\ \hat v \\ \hat p \end{pmatrix} = \mathcal{E} \begin{pmatrix} \hat\rho \\ \hat v \\ \hat p \end{pmatrix} \qquad \mathcal{E} = \begin{pmatrix} 0 & k - \mathbb{i}\mu & 0 \\ \mathbb{i}g & 0 & k - \mathbb{i}\mu \\ 0 & c^2k - \mathbb{i}c^2\mu & 0 \end{pmatrix} \tag{A.8}$$

The eigenvalues of $\mathcal{E}$ are now real: $\omega_1 = 0$, $\omega_{2,3} = \pm c\sqrt{k^2 + \mu^2}$. Although this transformation brings the endeavour of finding the exact solution to (3.6)–(3.8) into the realm of the possible, technical difficulties prevent one from actually computing all Green's functions in closed form.

Assume therefore that the only non-vanishing initial data are in the velocity. Then the Fourier mode at initial time reads

$$\begin{pmatrix} 0 \\ \hat v \\ 0 \end{pmatrix} \exp(\mathbb{i}kx) \tag{A.9}$$

and at a later time it becomes

$$\sum_{m=1}^{3} v_m \exp(\mathbb{i}kx - \mathbb{i}\omega_m t) \tag{A.10}$$

where the decomposition of $\begin{pmatrix} 0 \\ \hat{v} \\ 0 \end{pmatrix}$ in the eigenbasis of $\mathcal{E}$ is used, i.e.

$$\begin{pmatrix} 0 \\ \hat{v} \\ 0 \end{pmatrix} = \sum_{m=1}^{3} v_m \qquad\qquad \mathcal{E} v_m = \omega_m v_m \tag{A.11}$$

Such a basis is given e.g. by

$$e_1 = \begin{pmatrix} \mu + \mathbb{i}k \\ 0 \\ g \end{pmatrix} \qquad\qquad e_{2,3} = \begin{pmatrix} \mu + \mathbb{i}k \\ \pm \mathbb{i} c\sqrt{k^2+\mu^2} \\ c^2(\mu + \mathbb{i}k) \end{pmatrix} \tag{A.12}$$

Collecting the terms yields the time evolution of the Fourier mode (A.9):

$$\hat{v}\exp(\mathbb{i}kx) \begin{pmatrix} -\dfrac{(\mu+\mathbb{i}k)\sin\left(ct\sqrt{k^2+\mu^2}\right)}{c\sqrt{k^2+\mu^2}} \\ \cos\left(ct\sqrt{k^2+\mu^2}\right) \\ -\dfrac{c^2(\mu+\mathbb{i}k)\sin\left(ct\sqrt{k^2+\mu^2}\right)}{c\sqrt{k^2+\mu^2}} \end{pmatrix} \tag{A.13}$$

$$= \hat{v} \begin{pmatrix} -(\mu + \partial_x) \\ \partial_t \\ -c^2(\mu + \partial_x) \end{pmatrix} \exp(\mathbb{i}kx) \frac{\sin\left(ct\sqrt{k^2+\mu^2}\right)}{c\sqrt{k^2+\mu^2}} \tag{A.14}$$

Green's function is obtained by inserting the Fourier transform of a Dirac $\delta_{x'}$ at x', i.e. taking $\hat{v} = \frac{\exp(-\mathbb{i}kx')}{\sqrt{2\pi}}$ and performing the inverse Fourier transform with the help of formula 1.7 (30) in [Bat54]. This yields, wherever defined,

$$\begin{pmatrix} G_\rho(t,x;x') \\ G_v(t,x;x') \\ G_p(t,x;x') \end{pmatrix} = \begin{pmatrix} -(\mu + \partial_x) \\ \partial_t \\ -c^2(\mu + \partial_x) \end{pmatrix} \frac{1}{2c} J_0\left(\mu\sqrt{(ct)^2 - (x-x')^2}\right) \tag{A.15}$$

$$+ \begin{pmatrix} -\dfrac{\delta_{x+ct} - \delta_{x-ct}}{2c} \\ \dfrac{\delta_{x+ct} + \delta_{x-ct}}{2} \\ c\left(\delta_{x+ct} - \delta_{x-ct}\right) \end{pmatrix} \tag{A.16}$$

where J_0 is the 0-th order Bessel function of the first kind, and $J_0' = -J_1$. Then the solution is obtained by performing a convolution with the initial data. Reinstalling the tilde one has

$$\tilde{v}(t,x) = \int \mathrm{d}x'\, G_v(t,x;x')\tilde{v}_0(x') \tag{A.17}$$

$$v(t,x) = \int \mathrm{d}x'\, G_v(t,x;x')\mathrm{e}^{\mu(x-x')}v_0(x') \tag{A.18}$$

$$= \frac{1}{2}\int \mathrm{d}x'\, \mathrm{e}^{\mu(x-x')}\partial_{ct}J_0\left(\mu\sqrt{(ct)^2-(x-x')^2}\right)v_0(x') \tag{A.19}$$

$$+ \frac{1}{2}\Big(\mathrm{e}^{-\mu ct}v_0(x+ct) + \mathrm{e}^{\mu ct}v_0(x-ct)\Big) \tag{A.20}$$

$$\rho(t,x) = -\frac{1}{2c}\int \mathrm{d}x'\, \mathrm{e}^{\mu(x-x')}\left(\mu+\partial_x\right)J_0\left(\mu\sqrt{(ct)^2-(x-x')^2}\right)v_0(x') \tag{A.21}$$

$$- \frac{1}{2c}\Big(\mathrm{e}^{-\mu ct}v_0(x+ct) - \mathrm{e}^{\mu ct}v_0(x-ct)\Big) \tag{A.22}$$

and analogously for p. However, it is easier to note that

$$\partial_t(c^2\rho - p) = 0 \tag{A.23}$$

such that

$$p(t,x) = p_0(x) + c^2\Big(\rho(t,x) - \rho_0(x)\Big) \tag{A.24}$$

Acknowledgments. We thank Philip L. Roe for valuable comments and advice.

REFERENCES

[ABB+04] Emmanuel Audusse, François Bouchut, Marie-Odile Bristeau, Rupert Klein, and Benoit Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM Journal on Scientific Computing*, 25(6):2050–2065, 2004.

[Bar18] Wasilij Barsukow. *Low Mach number finite volume methods for the acoustic and Euler equations.* Doctoral thesis, University of Wuerzburg, 2018.

[Bar19a] W Barsukow. The active flux scheme for nonlinear problems which admit characteristic variables. *submitted to J. Sci. Comp*, 2019.

[Bar19b] Wasilij Barsukow. Stationarity preserving schemes for multi-dimensional linear systems. *Mathematics of Computation*, 88(318):1621–1645, 2019.

[Bat54] Harry Bateman. *Tables of integral transforms (volume 1)*, volume 1. McGraw-Hill Book Company, 1954.

[BCK16] Jonas P Berberich, Praveen Chandrashekar, and Christian Klingenberg. A general well-balanced finite volume scheme for euler equations with gravity. In *XVI International Conference on Hyperbolic Problems: Theory, Numerics, Applications*, pages 151–163. Springer, 2016.

[BCK19] Jonas P Berberich, Praveen Chandrashekar, and Christian Klingenberg. High order well-balanced finite volume methods for multi-dimensional systems of hyperbolic balance laws. *arXiv preprint arXiv:1903.05154*, 2019.

[BCKR19] Jonas P Berberich, Praveen Chandrashekar, Christian Klingenberg, and Friedrich K Röpke. Second order finite volume scheme for Euler equations with gravity which is well-balanced for general equations of state and grid systems. *Communications in Computational Physics*, 26:599–630, 2019.

[BHKR19] Wasilij Barsukow, Jonathan Hohm, Christian Klingenberg, and Philip L Roe. The active flux scheme on Cartesian grids and its low Mach number limit. *Journal of Scientific Computing*, 81(1):594–622, 2019.

[BKCK20] Jonas P Berberich, Roger Käppeli, Praveen Chandrashekar, and Christian Klingenberg. High order discretely well-balanced finite volume methods for Euler equations with gravity – without any a priori information about the hydrostatic solution. *arXiv preprint arXiv:2005.01811*, 2020.

[BV94] Alfredo Bermudez and Ma Elena Vázquez. Upwind methods for hyperbolic conservation laws with source terms. *Computers & Fluids*, 23(8):1049–1071, 1994.

[CCK+18] Alina Chertock, Shumo Cui, Alexander Kurganov, Şeyma Nur Özcan, and Eitan Tadmor. Well-balanced schemes for the Euler equations with gravitation: Conservative formulation using global fluxes. *Journal of Computational Physics*, 2018.

[CK15] Praveen Chandrashekar and Christian Klingenberg. A second order well-balanced finite volume scheme for Euler equations with gravity. *SIAM Journal on Scientific Computing*, 37(3):B382–B402, 2015.

[CL94] P Cargo and AY LeRoux. A well balanced scheme for a model of atmosphere with gravity. *COMPTES RENDUS DE L ACADEMIE DES SCIENCES SERIE I-MATHEMATIQUE*, 318(1):73–76, 1994.

[DZBK14] Vivien Desveaux, Markus Zenk, Christophe Berthon, and Christian Klingenberg. A well-balanced scheme for the Euler equation with a gravitational potential. In *Finite Volumes for Complex Applications VII-Methods and Theoretical Aspects*, pages 217–226. Springer, 2014.

[DZBK16] Vivien Desveaux, Markus Zenk, Christophe Berthon, and Christian Klingenberg. A well-balanced scheme to capture non-explicit steady states in the Euler equations with gravity. *International Journal for Numerical Methods in Fluids*, 81(2):104–127, 2016.

[ER13] Timothy A Eymann and Philip L Roe. Multidimensional active flux schemes. In *21st AIAA computational fluid dynamics conference*, 2013.

[Fan17] Duoming Fan. *On the acoustic component of active flux schemes for nonlinear hyperbolic conservation laws.* PhD thesis, University of Michigan, Dissertation, 2017.

[FR15] Doreen Fan and Philip L Roe. Investigations of a new scheme for wave propagation. In *22nd AIAA Computational Fluid Dynamics Conference*, page 2449, 2015.

[GL96] Joshua M Greenberg and Alain-Yves LeRoux. A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM Journal on Numerical Analysis*, 33(1):1–16, 1996.

[HKS19] Christiane Helzel, David Kerkmann, and Leonardo Scandurra. A new ADER method inspired by the active flux method. *Journal of Scientific Computing*, 80(3):1463–1497, 2019.

[KM16] R Käppeli and S Mishra. A well-balanced finite volume scheme for the Euler equations with gravitation-the exact preservation of hydrostatic equilibrium with arbitrary entropy stratification. *Astronomy & Astrophysics*, 587:A94, 2016.

[LeV98] Randall J LeVeque. Balancing source terms and flux gradients in high-resolution Godunov methods: the quasi-steady wave-propagation algorithm. *Journal of computational physics*, 146(1):346–365, 1998.

[LGB11] Randall J LeVeque, David L George, and Marsha J Berger. Tsunami modelling with adaptively refined finite volume methods. *Acta Numerica*, 20:211–289, 2011.

[NR16] Hiroaki Nishikawa and Philip L Roe. Third-order active-flux scheme for advection diffusion: hyperbolic diffusion, boundary condition, and Newton solver. *Computers & Fluids*, 125:71–81, 2016.

[VL77] Bram Van Leer. Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. *Journal of computational physics*, 23(3):276–299, 1977.