

ON THE ORIGIN OF DIVERGENCE ERRORS IN MHD SIMULATIONS AND CONSEQUENCES FOR NUMERICAL SCHEMES

FRIEDEMANN KEMM

ABSTRACT. This paper investigates the origin of divergence errors in MHD simulations. For that purpose, we introduce the concept of discrete involutions for discretized conservation laws. This is done in analogue to the concept of involutions for hyperbolic conservation laws, introduced by Dafermos. By exploring the connection between discrete involutions and resonance, especially for constrained transport like MHD, we identify the lack of positive central viscosity and the assumption of one-dimensional physics in the calculation of intercell fluxes as the main sources of divergence errors. As an example for the consequences for numerical schemes, we give a hint how to modify Roe-type schemes in order to decrease the divergence errors considerably and, thus, stabilize the scheme.

1. INTRODUCTION

Hyperbolic conservation laws are usually equipped with additional conditions. Most important is the existence of a convex entropy, which singles out the physical relevant solution out of the large set of possible weak solutions. Sometimes, especially when there is no convex entropy, or the system degenerates into a weakly or resonant hyperbolic system, other laws have to be included to find physical solutions. In the first case (convex entropy), the additional law is for an additional variable, namely the entropy, which depends on the state variables, but is no state variable itself. In the latter case, we have additional partial differential equations for the state variables themselves. In the first case the additional law is a partial differential equation or inequality of evolution type, usually a conservation law itself, in the second it is a first order non-evolutionary constraint. These additional constraints are, as Dafermos points out [10, 9], involutions for the underlying system of conservation laws. So the resulting system, which includes both, the evolution system and the condition, has more equations than unknowns. If the involution is satisfied by the initial state for the evolution equations, it is satisfied by the solution of the evolution system for all times. Thus, in the continuous setting, the constraint is merely a condition on the initial state.

These constraints play an important role in many branches of physics, the most famous of which is the area of electromagnetic modelling and plasma physics. Here we face constraints for the electric field as well as for the magnetic field. In numerical simulations this may cause severe problems, because in general it is impossible to reproduce these conditions exactly. This results in unphysical forces and thus completely useless solutions [6, 31], especially in Magnetohydrodynamics (MHD). In MHD many codes completely fail. But this is not the case with all numerical schemes.

2000 *Mathematics Subject Classification.* Primary 76W05 39A12, 35L45, 35L65, 35L80; Secondary 35N10, 65M06, 39A70, 65Z05,

Key words and phrases. involutions, constraint, magnetohydrodynamics, plasma physics, Maxwell equations, divergence, curl, operator scheme, finite differences, finite volume method, resonance, hyperbolic PDE, compressible flow.

First, there are approaches, which are designed to model the constraint numerically. Many of them are done on staggered grids [3, 4, 13]. Some newer approaches also work on collocated grids [39, 38, 35, 26, 27, 28, 15, 14, 40] or in the context of Discontinuous-Galerkin schemes [5]. Usually this class of schemes is referred to as *constrained transport*.

A second family of schemes are based on a modification of the system of partial differential equations which makes the constraint part of the evolution system itself. In the context of plasma physics, a popular approach is to transport the involution term, in this case the divergence of the magnetic field, with the flow velocity. This was first suggested for numerical simulations by Brackbill and Barnes [6] and put forward by employing the full symmetrizable form of the MHD-equations by Powell et al. [33, 34]. In [15, 14, 40], this approach is even combined with constrained transport. Another possibility is to apply a kind of a generalized Lagrange multiplier approach [31], a method which can show up in several variants: resulting in a Hodge-projection scheme, resulting in a parabolic treatment of the involution term [25], resulting in a hyperbolic system – the involution term is radiated with an artificial speed out of the computational domain [30, 29] – or it results in a treatment of the involution in the manner of a telegraph equation [11, 23]. In the context of electromagnetic models and plasma physics these approaches are usually referred to as *divergence cleaning*.

A third class of schemes are stable without any modification or special discretization technique. This is the case in many physical contexts. For Magnetohydrodynamics it is only reported very scarcely. But still there are some examples: The scheme of Zachary, Malagoli and Colella [41], an upwind method, published already in 1994, has this property. Another example is the scheme presented by Balbás and Tadmor[2], which in contrary is a central scheme. They still need some intermediate cleaning steps to obtain physical relevant solutions, but only at a few time steps in a long interval [41, 1]. Both schemes have in common that they discretize the full equations, while most schemes, for the computation of inter-cell fluxes, employ one-dimensional physics in the direction of the normal of the cell face. This shows that there is something special in discretizing the multidimensional equations directly.

Therefore, in this study, we put forward the investigations started in [21] and take a closer look at involutions and their relation to constrained transport and resonance. We also take a closer look at the discretization of conservation laws in terms of finite differences for the partial derivatives in the equations. We define discrete analogues of the most important types of involutions and look for sufficient conditions for the existence of discrete involutions to a given discretized conservation law. We single out a class of linear schemes for which the discrete involutions are exact. We consider the interplay between discrete involutions and resonance, and we study the role of central numerical viscosity and the assumption of one-dimensional physics in the computation of inter-cell fluxes. As a result, applying full physics in the computation of inter-cell fluxes and a suitable amount of central viscosity on the resonant wave lead to a stable scheme also in the MHD context. As an example, we show how to apply the Harten entropy-fix in a smart way to adjust the viscosity on the resonant wave. There are still some divergence errors, but the work needed in divergence cleaning can be considerably reduced.

The plan of the paper is as follows: First, we give an overview on the concept of involutions and its connections to resonance and constrained transport. Section 3 presents a theory for discrete involutions. Also some standard schemes are investigated whether they yield exact or only approximate involutions. The key is a shift in the interpretation

of numerical schemes. Some terms, traditionally considered to be part of the spatial discretization, are identified to be essentially part of the time discretization. This helps us in Sect. 4 to investigate the interplay between discrete involutions, resonance, central numerical viscosity, and the assumption of one-dimensional physics. Also we show numerical evidence of the theoretical results (Section 4.4). In this course, we present, as an example, a modification of the Roe-scheme which minimizes the production of divergence errors.

2. HYPERBOLIC CONSERVATION LAWS WITH INVOLUTIONS

2.1. Definition and a sufficient condition. The starting point of our considerations is the general conservation law

$$(1) \quad \mathbf{q}_t + \nabla \cdot \mathbf{F}(\mathbf{q}) = 0 ,$$

where \mathbf{q} denotes the vector of conserved quantities and $\mathbf{F} = (F_1, F_2, \dots)$ denotes the flux. The F_i are the directional fluxes in the (space) directions given by the standard unit vectors \mathbf{e}_i . The corresponding flux Jacobians will be denoted by \mathbf{A}_i .

The system (1) is called *hyperbolic* if for all directions

$$\mathbf{n} = \sum_i n_i \mathbf{e}_i , \quad \|\mathbf{n}\| = 1 ,$$

the according flux Jacobian

$$\mathbf{A}_{\mathbf{n}} = \sum_i n_i \mathbf{A}_i$$

is diagonalizable with real eigenvalues. It is called *strictly hyperbolic* if, in addition, all eigenvalues are distinct. If it is not diagonalizable but still all eigenvalues are real it is called *weakly hyperbolic* or *resonant hyperbolic*. In this survey we restrict the analysis to systems which are, at least, weakly hyperbolic.

The system (1) has an *involution* if there exist constant matrices \mathbf{M}_i so that the condition

$$(2) \quad \sum_i \mathbf{M}_i \mathbf{q}_{x_i} = 0 ,$$

also called the involution of system (1), holds true for all times if it is satisfied by the initial data.

In his work on hyperbolic systems with involutions, Dafermos [9, 10] concentrates on a subclass which includes most of the physically relevant cases:

Theorem 1. *Let the system (1) and matrices \mathbf{M}_i be given. If the directional fluxes fulfil the following antisymmetric condition*

$$(3) \quad \mathbf{M}_i F_j + \mathbf{M}_j F_i = 0 \quad \forall i, j = 1, 2, \dots ,$$

then

$$\sum_i \mathbf{M}_i \mathbf{q}_{x_i}$$

is an involution of system (1) and satisfies the additional condition

$$(4) \quad \frac{\partial}{\partial t} \left(\sum_i \mathbf{M}_i \mathbf{q}_{x_i} \right) = 0 .$$

As a consequence, $\sum_i \mathbf{M}_i \mathbf{q}_{x_i}$ not only makes up an involution of (1), but in addition is constant in time, independent of the initial state. This shows that (3) is a rather strong condition. But, as Sect. 2.2 shows, many important systems satisfy condition (3). In the following sections this is used to find discrete analogues for the concept of involutions.

For the understanding of the following sections, it is necessary to understand the proof of Theorem 1. Summed up the proof consists in four steps:

- (1) Apply $\sum_i \mathbf{M}_i \frac{\partial}{\partial x_i}$ to the conservation law (1).
- (2) Constant matrices commute with partial derivatives.
- (3) Partial derivatives commute with each other.
- (4) Due to condition (3) all terms including fluxes vanish.

In more detail, we find after application of step 1

$$\sum_i \mathbf{M}_i \frac{\partial}{\partial x_i} \frac{\partial}{\partial t} \mathbf{q} + \sum_i \mathbf{M}_i \frac{\partial}{\partial x_i} \sum_j \frac{\partial}{\partial x_j} F_j(\mathbf{q}) = 0 .$$

Now we make use of the facts 2 and 3 to obtain

$$\frac{\partial}{\partial t} \left(\sum_i \mathbf{M}_i \frac{\partial}{\partial x_i} \mathbf{q} \right) + \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \mathbf{M}_i F_j(\mathbf{q}) = 0 .$$

Since the operator $\frac{\partial^2}{\partial x_i \partial x_j}$ is symmetric and, according to (3), $\mathbf{M}_i F_j(\mathbf{q})$ is antisymmetric, the last sum vanishes, which completes the proof. \square

2.2. Examples for hyperbolic systems with involutions. A whole bunch of examples for systems with involutions satisfying (3) can be found in the above mentioned works by Dafermos [9, 10] and in the studies by Torrilhon and Fey [38, 39, 37]. Here we present only few of them:

As an introductory example, Dafermos [10] presents the equations for isentropic processes of thermoelastic nonconductors of heat:

$$(5) \quad \begin{aligned} \mathbf{F}_t - \nabla \mathbf{v} &= 0 , \\ \mathbf{v}_t - \nabla \cdot \mathbf{T}(\mathbf{F}) &= 0 \end{aligned}$$

with the deformation gradient \mathbf{F} , velocity \mathbf{v} , and the stress tensor \mathbf{T} . Since the time evolution of the deformation gradient \mathbf{F} is a gradient, it is curl-free. Therefore $\nabla \times \mathbf{F}$ is an involution for system (5). The matrices \mathbf{M}_i ($i = 1, \dots, 3$) are in $\mathbb{R}^{3 \times 6}$, and, while the right half is zero, the left half reads as

$$(6) \quad \mathbf{M}_1^{\text{left}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \mathbf{M}_2^{\text{left}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \mathbf{M}_3^{\text{left}} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

With these matrices, condition (3) is satisfied.

An important hyperbolic system with involution are the vacuum Maxwell equations

$$(7) \quad \mathbf{E}_t - c^2(\nabla \times \mathbf{B}) = -\frac{\mathbf{j}}{\varepsilon_0} ,$$

$$(8) \quad \mathbf{B}_t + (\nabla \times \mathbf{E}) = 0 ,$$

$$(9) \quad \nabla \cdot \mathbf{E} = \frac{q}{\varepsilon_0} ,$$

$$(10) \quad \nabla \cdot \mathbf{B} = 0$$

with the electric field \mathbf{E} , magnetic induction \mathbf{B} , electric current \mathbf{j} , charge density q , speed of light c , and the constants ε_0 and μ_0 . In the absence of electric charge and current this is a homogeneous hyperbolic conservation law, where the divergence of both fields, \mathbf{E} and \mathbf{B} is preserved. The matrices involved in condition (2) and (3) are

$$(11) \quad \mathbf{M}_i = \begin{pmatrix} \mathbf{e}_i^T & \mathbf{0}^T \\ \mathbf{0}^T & \mathbf{e}_i^T \end{pmatrix}.$$

Since the models of plasma-physics are obtained by using the Maxwell equations, they also inherit the involutions. In the MHD-equations no evolution equation for the electric field is included. Thus, only the divergence of the magnetic field is inherited as an involution. The full equations of ideal compressible magnetohydrodynamics are

$$(12) \quad \rho_t + \nabla \cdot [\rho \mathbf{v}] = 0 ,$$

$$(13) \quad (\rho \mathbf{v})_t + \nabla \cdot [\rho \mathbf{v} \circ \mathbf{v} + (p + \frac{1}{2} \mathbf{B}^2) \mathbf{I} - \mathbf{B} \circ \mathbf{B}] = 0 ,$$

$$(14) \quad \mathbf{B}_t + \nabla \cdot [\mathbf{B} \circ \mathbf{v} - \mathbf{v} \circ \mathbf{B}] = 0 ,$$

$$(15) \quad e_t + \nabla \cdot [(e + p + \frac{1}{2} \mathbf{B}^2) \mathbf{v} - \mathbf{B}(\mathbf{v} \cdot \mathbf{B})] = 0 ,$$

$$(16) \quad \nabla \cdot \mathbf{B} = 0 .$$

The last equation (16) is the involution for the evolution system (12)–(15). The asymmetric condition (3) is satisfied with $\mathbf{M}_i = (0, 0, 0, 0, \mathbf{e}_i^T, 0)$. Thus, the MHD-equations nicely fit into the framework given by Dafermos [9, 10].

In [16], Gilman argues that the classical “shallow water” equations of geophysical fluid dynamics should be useful for studying the global dynamics of the solar tachocline and demonstrates the existence of an MHD analogon that would allow taking into account the strong toroidal magnetic field likely to be present there. So he presents a derivation analogous to that for the classical shallow water equations and comes up with the following system of shallow water magnetohydrodynamics (SMHD)

$$(17) \quad \begin{aligned} h_t + \nabla \cdot [h \mathbf{v}] &= 0 , \\ (h \mathbf{v})_t + \nabla \cdot [h \mathbf{v} \circ \mathbf{v} - h \mathbf{B} \circ \mathbf{B} + \frac{gh^2}{2} \mathbf{I}] &= \mathbf{0} , \\ (h \mathbf{B})_t - \nabla \times [\mathbf{v} \times (h \mathbf{B})] &= \mathbf{0} \end{aligned}$$

with the involution

$$(18) \quad \nabla \cdot (h \mathbf{B}) = 0 ,$$

This inherits most of its behaviour from the original MHD-system (12)–(15). The main difference is that, due to the averaging over the third space dimension, the magnetic field \mathbf{B} is now replaced by $h \mathbf{B}$, where h denotes the height of the fluid layer under consideration and g is the gravitational constant.

Since the structure of the critical part of the evolution for the vorticity transport (??), MHD, and SMHD is similar, we also consider the linear model problem of Fey and Torrilhon [39], which resembles the common behaviour of those systems in the simplest possible setting. For a given velocity field \mathbf{v} , constant in space and time, we consider

$$(19) \quad \mathbf{B}_t - \nabla \times (\mathbf{v} \times \mathbf{B}) = \mathbf{0} ,$$

or in divergence form,

$$(20) \quad \mathbf{B}_t + \nabla \cdot (\mathbf{B} \circ \mathbf{v} - \mathbf{v} \circ \mathbf{B}) = \mathbf{0} .$$

Obviously the asymmetric condition (3) is satisfied with $\mathbf{M}_i = \mathbf{e}_i^T$ and the divergence of \mathbf{B} makes up an involution for the system. This is a model for divergence-preserving transport.

For the sake of completeness, we also present the model for curl-preserving transport given by Fey and Torrilhon [39]

$$(21) \quad \mathbf{P}_t + \nabla(\mathbf{v} \cdot \mathbf{P}) = \mathbf{0} ,$$

or in divergence form,

$$(22) \quad \mathbf{P}_t + \nabla \cdot ([\mathbf{v} \cdot \mathbf{P}] \mathbf{I}) = \mathbf{0} .$$

Here again, it can be seen from the flux form that (3) is satisfied. The matrices \mathbf{M}_i are the same as those presented in (6), and from (21) it is seen that the resulting involution is $\nabla \times P$.

The last four systems, MHD, SMHD, and the model systems for constraint preserving transport, have one point in common: dependent on the velocity field, they might lose full hyperbolicity. In general they are only weakly, or resonant, hyperbolic.

2.3. Resonant hyperbolic problems and involutions. In this section we consider the relation between involutions and resonant hyperbolic systems.

2.3.1. *Resonance.* In physics, systems which allow for solutions growing unboundedly in time, usually are called resonant. The most famous example is the harmonic oscillator with a periodic exciting force. If the frequency of the excitation meets the eigenfrequency of the system, the amplitude grows unboundedly. A similar behaviour can be found for weakly hyperbolic systems. The model equations for divergence-preserving transport (19) provide a nice example for resonance. Let in the 2-d case $\mathbf{v} = (u, v)^T = (0, v)^T$. Then the system is only weakly hyperbolic and reads as

$$(23) \quad B_{1t} + vB_{1y} = 0 ,$$

$$(24) \quad B_{2t} - vB_{1x} = 0 .$$

This means that B_1 is transported in y -direction and acts as a source for B_2 . If B_1 varies only in y -direction, the source is constant and, thus B_2 grows unboundedly with a constant rate. We will go back to this example later.

Of the systems with evolutions provided in the previous section, some are fully hyperbolic, some are only resonant hyperbolic. Dafermos [10] points out that system (5) is hyperbolic if the inner energy, which defines the stress tensor, is rank-one convex. Thus, hyperbolicity depends on the state.

Although in any space direction all wave speeds are $\pm c$, the Maxwell equations are fully hyperbolic. They allow for no resonant effects except from those introduced by outer source terms.

In contrast, the MHD and SMHD equations allow for resonant states. By using the magnetohydrodynamic approximation for the electric field, $\mathbf{E} \approx -\mathbf{v} \times \mathbf{B}$, the induction equation attains the structure of divergence preserving transport. If we set $\mathbf{B} = (B_1 = 0, B_2, B_3)^T$, $\mathbf{v} = (u = 0, v, w)^T$, i. e. velocity and magnetic field are perpendicular to the first space direction, then the flux Jacobian in that direction has zero as a sixfold eigenvector with five-dimensional eigenspace. The system is only resonant hyperbolic. Whenever the velocity and the magnetic field are in one plane, the flux Jacobian in the direction perpendicular to that plane is deficient, the system is only resonant hyperbolic. A similar situation occurs when the velocity component parallel to the magnetic field equals $\pm a$, where a is the speed of sound, and the magnetic field is $a\sqrt{\rho}$. Then zero is a fourfold eigenvalue with three-dimensional eigenspace. Again the system is only resonant hyperbolic.

Due to the reduction of the physical problem to two space dimensions, for the shallow water MHD the same resonance as for MHD occurs when velocity and magnetic field are parallel. Another resonant case can be found if we have $\mathbf{v} = (u = \pm c_g, v)^T$, where $c_g = \sqrt{B_1^2 + gh}$ is the magneto-gravitational speed. In this case, zero is a double eigenvalue with one-dimensional eigenspace.

The model system for divergence preserving transport (19) shares the resonant behaviour, as can be seen at the beginning of this section. The flux Jacobian in the direction perpendicular to the velocity is deficient, the system is only resonant hyperbolic. For the model system for curl-preserving transport, the situation is similar.

We will go into detail for this in the following sections.

2.4. Relation of involutions to zero eigenvalues and resonance. In this section we investigate the connection between involutions, zero eigenvalues and resonance in more detail. First we want to recall some considerations of Dafermos [10]. The antisymmetric condition (3) is equivalent to

$$(25) \quad \mathbf{M}_i \mathbf{A}_j + \mathbf{M}_j \mathbf{A}_i = \mathbf{0} \quad \forall i, j.$$

If we take the unit vector $\mathbf{n} = (n_1, n_2, n_3)^T$ we find for the flux Jacobian $\mathbf{A}_{\mathbf{n}}$ in direction of \mathbf{n}

$$(26) \quad \begin{aligned} \mathbf{M}_{\mathbf{n}} \mathbf{A}_{\mathbf{n}} &= \left(\sum_i n_i \mathbf{M}_i \right) \left(\sum_j n_j \mathbf{A}_j \right) \\ &= \sum_{i,j} n_i n_j \mathbf{M}_i \mathbf{A}_j \\ &= \mathbf{0}. \end{aligned}$$

As a consequence, the range of $\mathbf{A}_{\mathbf{n}}$ is a subset of the kernel of $\mathbf{M}_{\mathbf{n}}$, and therefore the dimension of the kernel of $\mathbf{A}_{\mathbf{n}}$ is greater than or equal to the rank of $\mathbf{M}_{\mathbf{n}}$. In particular, we know that it is at least one. We always have a zero eigenvalue for systems which satisfy the antisymmetric condition (3), and equality would mean that the rows of $\mathbf{M}_{\mathbf{n}}$ are just the left eigenvectors of $\mathbf{A}_{\mathbf{n}}$ for the zero eigenvalue. As a consequence, in the case of equality, the zero eigenvalue has a full set of eigenvectors and, thus, can not destroy the hyperbolicity of the system. An example for this are the vacuum Maxwell equations with zero as a double eigenvalue and we have

$$(27) \quad \mathbf{M}_i = \begin{pmatrix} \mathbf{e}_i^T & \mathbf{0}^T \\ \mathbf{0}^T & \mathbf{e}_i^T \end{pmatrix},$$

which makes up a linearly independent set of two left eigenvectors. In the case that the range of $\mathbf{A}_{\mathbf{n}}$ is a proper subset of the kernel of $\mathbf{M}_{\mathbf{n}}$ things might be worse, as can be seen with the above example systems.

If we apply these considerations to the divergence-free transport (19), we find that the rank of $\mathbf{M}_{\mathbf{n}}$ is one for all directions \mathbf{n} . But if $\mathbf{n} \perp \mathbf{v}$, there is no transport in direction of \mathbf{n} , and hence, the multiplicity of the zero eigenvalue of $\mathbf{A}_{\mathbf{n}}$ is three. If \mathbf{v} , and thus also $\mathbf{A}_{\mathbf{n}}$, does not vanish completely the system matrix can not be diagonalized, because the eigenspace has dimension two. A similar consideration holds for the curl-free transport (21). In this case, except for $\mathbf{v} = \mathbf{0}$, the rows of $\mathbf{M}_{\mathbf{n}}$ always form a basis of the space of left eigenvectors of $\mathbf{A}_{\mathbf{n}}$, proving again that both prototypes for constrained transport are merely resonant hyperbolic.

2.4.1. A quantitative view on resonance for divergence- and curl-preserving transport.

For a quantitative view on resonance for divergence-preserving transport, we revisit the example (29), (30) from the beginning of the last section and look at it in more detail: let in the 2-d case $\mathbf{v} = (u, v)^T = (0, v)^T$. It follows from the considerations at the beginning of Sect. 2.4 that \mathbf{A}_1 is not diagonalizable. We can therefore expect resonance phenomena in the first space direction. Since $u = 0$ and

$$(28) \quad \mathbf{A}_{\mathbf{n}} = (\mathbf{n}^T \cdot \mathbf{v}) \mathbf{I} - \mathbf{v} \circ \mathbf{n},$$

we can rewrite the system as

$$(29) \quad B_{1t} + vB_{1y} = 0 ,$$

$$(30) \quad B_{2t} + vB_{2y} = v(B_{1x} + B_{2y}) = v(\nabla \cdot \mathbf{B}) .$$

The source is in the evolution equation of the second component of \mathbf{B} and is proportional to the involution $\nabla \cdot \mathbf{B}$. If \mathbf{B} is divergence-free there is no resonance at all. In general the 2-d system can be rewritten as

$$(31) \quad \mathbf{B}_t + \sum_i v_i \mathbf{B}_{x_i} = - \left(\sum_j \mathbf{A}_j \mathbf{M}_j^T \right) \sum_i \mathbf{M}_i \mathbf{B}_{x_i} .$$

Investigating the right hand side of this equation, we find just the negative of the Powell correction term [33]. Therefore, if we had added the Powell correction term to the right hand side of system (19), we would have obtained a non resonant, fully hyperbolic system, in this simple linear case pure advection. Although not pure advection, the Powell system in the full MHD equations is also fully hyperbolic without any resonance.

In the 3-d case there is just one additional factor to include. The system can be rewritten as

$$(32) \quad \mathbf{B}_t + \sum_i v_i \mathbf{B}_{x_i} = - \frac{1}{2} \left(\sum_j \mathbf{A}_j \mathbf{M}_j^T \right) \sum_i \mathbf{M}_i \mathbf{B}_{x_i} .$$

Adding $\frac{1}{2} \left(\sum_j \mathbf{A}_j \mathbf{M}_j^T \right) \sum_i \mathbf{M}_i \mathbf{B}_{x_i}$ to the right hand side of system (19) would lead to pure advection and, thus, to a fully hyperbolic system.

In full MHD in three space dimensions with the usual ordering of the equations, the addition of $\frac{1}{2} \left(\sum_j \mathbf{A}_j \mathbf{M}_j^T \right) \sum_i \mathbf{M}_i \mathbf{B}_{x_i}$ with $\mathbf{M}_i = \mathbf{e}_{i+4}$, makes the resulting system fully hyperbolic. Nevertheless, for the use in numerical schemes, the original Powell correction is more convenient due to its simpler form of left and right eigenvectors.

In the same way, for curl-preserving transport (21), we get

$$(33) \quad \mathbf{P}_t + \sum_i v_i \mathbf{P}_{x_i} = - \left(\sum_j \mathbf{A}_j \mathbf{M}_j^T \right) \left(\sum_i \mathbf{M}_i \mathbf{P}_{x_i} \right) ,$$

where the last sum is just the involution. If the constraint is satisfied for the initial data, curl-preserving transport reduces to pure transport. Otherwise it is a transport with a source which is a linear function of the involution term. Since the involution term is constant in time, the source term is also constant in time. All in all the situation is quite similar to that in divergence preserving transport. Because of that, and because curl-preserving transport plays a minor part in practical applications, we won't go into further detail for that.

3. A DISCRETE ANALOGON TO THE CONCEPT OF INVOLUTIONS

This section is dedicated to the construction of discrete analogues of the concept of involutions for discretized conservation laws as well as a discrete analogue of Theorem 1 and its proof.

For this purpose, we first give some remarks on the notion and notation of Finite Difference schemes for hyperbolic conservation laws. This is necessary since the usual notation doesn't allow to transfer the results of Sect. 2.1 to the discrete case.

After that, we show how this transfer could be accomplished. We give discrete versions of Theorem 1 for semidiscrete, fully discrete, and a special case of linear schemes. In this context, we have to introduce exact and approximate discrete involutions.

Finally, we investigate some standard schemes. Which discrete version of Theorem 1 will apply to them? Will we find exact or only approximate discrete involutions? Although

rarely used in practice, these schemes are role models for most of the usual schemes, showing which behavior we have to expect from these methods.

3.1. On the notion and notation of Finite Difference schemes for hyperbolic conservation laws. In this paper we employ a rather strict, but still general, notion and notation of Finite Difference schemes for hyperbolic conservation laws. If \mathbf{I} is the set of all index vectors \mathbf{i} involved in the computation, including both, time- and space-indexes, a difference operator for some time-derivative is given by

$$(34) \quad \frac{\hat{\partial}}{\hat{\partial}t} \mathbf{q}_j = \frac{1}{\Delta t} \sum_{\mathbf{i} \in \mathbf{I}} \alpha_{\mathbf{i},j} \mathbf{q}_i ,$$

where the coefficients $\alpha_{\mathbf{i},j}$ are allowed to be matrix valued and to depend on anything, they only have to be bounded in time and space, and Δt is some characteristic time step size. We use the hat to distinguish difference operators from the corresponding derivatives. The inclusion of Δt into the formula makes the further considerations more convenient. For space derivatives we write in the same way

$$(35) \quad \frac{\hat{\partial}}{\hat{\partial}x_k} \mathbf{q}_j = \frac{1}{\Delta x} \sum_{\mathbf{i} \in \mathbf{I}} \beta_{\mathbf{i},j}^k \mathbf{q}_i ,$$

where the index k denotes the space direction, and Δx is some characteristic space step size, for example the minimal inradius of the grid cells. All other differential operators, like divergence, curl, gradient, mixed or higher derivatives, are discretized by means of the operators given in (34) and (35), where the difference operator for each space direction and for the time are fixed. Thus, for example, the second derivative of some quantity \mathbf{q} with respect to direction x_k has to be discretized by $\frac{\hat{\partial}}{\hat{\partial}x_k} \left(\frac{\hat{\partial}}{\hat{\partial}x_k} \mathbf{q} \right)$. We introduce this strict notation to be able to transfer the proof of Theorem 1 to the discrete case. As a consequence of the notation, in the following, the term $\hat{\partial}/\hat{\partial}t$ is merely an abbreviation for any discrete time difference of order q . This can be done because the difference between any two difference operators of order q is also $\mathcal{O}(\Delta t^q)$. For the other partial derivatives a similar consideration holds. This is a fact which we extensively use in our arguments. With these operators a discretized hyperbolic conservation law reads as:

$$(36) \quad \frac{\hat{\partial}}{\hat{\partial}t} \mathbf{q}_j + \sum_r \frac{\hat{\partial}}{\hat{\partial}x_r} \mathbf{F}_r(\mathbf{q}_j) = \mathbf{0} .$$

Note that this is not the way, the scheme is constructed. But any Finite Volume or Finite Difference scheme can artificially be rewritten in that way. This also is not the usual notation of discretized hyperbolic conservation laws in the literature. Normally, a simpler difference operator is chosen and applied to a system, where the physical flux function \mathbf{F} is replaced by a numeric flux function \mathbf{G} , which depends on the state in several grid cells. But this is not suitable for the investigation of discrete involutions, since we have to rely on the antisymmetric condition (3), which depends on \mathbf{F} and is usually not satisfied if \mathbf{F} is replaced by \mathbf{G} ¹. As we will see in Sect. 3.2.2, sometimes parts of \mathbf{G} have to be considered as a contribution to the discrete time derivative instead of the space derivative.

¹By applying the concept of numerical flux functions on a one-dimensional equidistant grid, it would be even possible to write all schemes, including implicit schemes, as $\frac{q_i^{n+1} - q_i^n}{\Delta t} - \frac{G_{i+1/2}^n - G_{i-1/2}^n}{\Delta x}$. All details are hidden in the definition of the numerical flux function G . In the same way, for every computational grid, a difference formulation can be found which only depends on the grid itself.

3.2. Proofs for discrete involutions. To prove the existence of discrete involutions, we first have to define them:

Definition 1. *If for a discretized hyperbolic conservation law (39)*

$$(37) \quad \frac{\hat{\partial}}{\hat{\partial}t} \left(\sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\hat{\partial}x_{\ell}} \mathbf{q}_{\mathbf{j}} \right) \rightarrow 0$$

as the time and space step size go to zero, it is called an approximate discrete involution for (39). If we have equality, i. e.

$$(38) \quad \frac{\hat{\partial}}{\hat{\partial}t} \left(\sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\hat{\partial}x_{\ell}} \mathbf{q}_{\mathbf{j}} \right) = 0 ,$$

we call it an exact discrete involution for (39).

We prove three discrete versions of Theorem 1, a general fully discrete, a semi-discrete version, and a linear special case.

3.2.1. The general fully discrete case. We start with the fully discrete general case. To prove that the antisymmetric condition (3) is sufficient for the existence of discrete involutions, we first have to investigate the commutators of the difference operators given in the previous section.

If we have for some quantity \mathbf{h}

$$(39) \quad \frac{\hat{\partial}}{\hat{\partial}x} \mathbf{h}_{\mathbf{j}} = (\mathbf{h}_x)_{\mathbf{j}} + \mathcal{O}(\Delta x^p)$$

and

$$(40) \quad \frac{\hat{\partial}}{\hat{\partial}t} \mathbf{h}_{\mathbf{j}} = \frac{1}{\Delta t} \sum_{\mathbf{i} \in \mathbf{I}} \alpha_{\mathbf{i}, \mathbf{j}} \mathbf{h}_{\mathbf{i}}$$

$$(41) \quad = (\mathbf{h}_t)_{\mathbf{j}} + \mathcal{O}(\Delta t^q) ,$$

we can verify the following:

$$(42) \quad \begin{aligned} \frac{\hat{\partial}}{\hat{\partial}t} \left(\frac{\hat{\partial}}{\hat{\partial}x} \mathbf{h} \right)_{\mathbf{j}} &= \frac{1}{\Delta t} \sum_{\mathbf{i} \in \mathbf{I}} \alpha_{\mathbf{i}, \mathbf{j}} [(\mathbf{h}_x)_{\mathbf{i}} + \mathcal{O}(\Delta x^p)] \\ &= \frac{\hat{\partial}}{\hat{\partial}t} (\mathbf{h}_x)_{\mathbf{j}} + \mathcal{O}\left(\frac{\Delta x^p}{\Delta t}\right) \\ &= (\mathbf{h}_{xt})_{\mathbf{j}} + \mathcal{O}(\Delta t^q) + \mathcal{O}\left(\frac{\Delta x^p}{\Delta t}\right) . \end{aligned}$$

With a similar consideration, for $\frac{\hat{\partial}}{\hat{\partial}x} \left(\frac{\hat{\partial}}{\hat{\partial}t} \mathbf{h} \right)_{\mathbf{j}}$, we find for the commutator of both discrete partial derivatives

$$(43) \quad \frac{\hat{\partial}}{\hat{\partial}t} \left(\frac{\hat{\partial}}{\hat{\partial}x} \mathbf{h} \right)_{\mathbf{j}} - \frac{\hat{\partial}}{\hat{\partial}x} \left(\frac{\hat{\partial}}{\hat{\partial}t} \mathbf{h} \right)_{\mathbf{j}} = \mathcal{O}\left(\frac{\Delta x^p}{\Delta t}\right) + \mathcal{O}\left(\frac{\Delta t^q}{\Delta x}\right) .$$

If for a simulation the time step and space step stay of the same order for all time, i. e. $\Delta t = \mathcal{O}_s(\Delta x)$ (the subscript s means that the order relation between Δt and Δx is symmetric), the commutator (43) simplifies to

$$(44) \quad \frac{\hat{\partial}}{\hat{\partial}t} \left(\frac{\hat{\partial}}{\hat{\partial}x} \mathbf{h} \right)_{\mathbf{j}} - \frac{\hat{\partial}}{\hat{\partial}x} \left(\frac{\hat{\partial}}{\hat{\partial}t} \mathbf{h} \right)_{\mathbf{j}} = \mathcal{O}(\Delta x^{\min\{p, q\} - 1}) .$$

But this is not always true, especially when resonance comes into play. If we consider the commutator of two different discrete space derivatives, e. g. in x - and y -direction, and both are of the same order of accuracy, p , we obtain

$$(45) \quad \frac{\hat{\partial}}{\hat{\partial}y} \left(\frac{\hat{\partial}}{\hat{\partial}x} \mathbf{h} \right)_j - \frac{\hat{\partial}}{\hat{\partial}x} \left(\frac{\hat{\partial}}{\hat{\partial}y} \mathbf{h} \right)_j = O(\Delta x^{p-1}).$$

The commutator of a discrete derivative and a Matrix \mathbf{M} can be obtained in the same way. It is

$$(46) \quad \mathbf{M} \frac{\hat{\partial}}{\hat{\partial}t} (\mathbf{h})_j - \frac{\hat{\partial}}{\hat{\partial}t} (\mathbf{M}\mathbf{h})_j = O(\Delta t^q).$$

Thus, no loss of accuracy is introduced.

With these preparations, the following Theorem can be proved:

Theorem 2. *Let the following (weakly) hyperbolic conservation law be given*

$$(47) \quad \mathbf{q}_t + \nabla \cdot \mathbf{F}(\mathbf{q}) = \mathbf{0}$$

together with constant matrices \mathbf{M}_ℓ which satisfy

$$(48) \quad \mathbf{M}_\ell F_r + \mathbf{M}_r F_\ell = \mathbf{0} \quad \forall \ell, r = 1, 2, \dots$$

Furthermore let $\hat{\partial}/\hat{\partial}t$ be a time discretization of order q and $\hat{\partial}/\hat{\partial}x_r$ be space differences of order p .

If we discretize the conservation system (47) with these discrete operators, then we obtain the following analogue of (4):

$$(49) \quad \frac{\hat{\partial}}{\hat{\partial}t} \left(\sum_\ell \mathbf{M}_\ell \frac{\hat{\partial}}{\hat{\partial}x_\ell} \mathbf{q}_j \right) = O\left(\frac{\Delta x^p}{\Delta t}\right) + O\left(\frac{\Delta t^q}{\Delta x}\right) + O(\Delta x^{p-1}) \quad \forall \mathbf{j}.$$

As a direct consequence, we can state the following corollary:

Corollary 1. *If in addition to the conditions of Theorem 2 the time and space step are of the same order, i. e. $\Delta t = O(\Delta x)$ and $\Delta x = O(\Delta t)$, then (49) can be simplified to*

$$(50) \quad \frac{\hat{\partial}}{\hat{\partial}t} \left(\sum_\ell \mathbf{M}_\ell \frac{\hat{\partial}}{\hat{\partial}x_\ell} \mathbf{q}_j \right) = O(\Delta x^{\min\{p,q\}-1}) \quad \forall \mathbf{j}.$$

This applies to linear systems and, in general, to nonlinear non-resonant systems. For general nonlinear resonant systems things might be worse. We will consider the general case in more detail in Sect. 4.

Proof. (of Theorem 2)

For a fixed index \mathbf{j} , the discretized conservation law reads

$$(51) \quad \frac{\hat{\partial}}{\hat{\partial}t} \mathbf{q}_j + \sum_r \frac{\hat{\partial}}{\hat{\partial}x_r} \mathbf{F}_r(\mathbf{q}_j) = \mathbf{0}.$$

Now we apply $\sum_\ell \mathbf{M}_\ell \frac{\hat{\partial}}{\hat{\partial}x_\ell}$ to that system:

$$(52) \quad \sum_\ell \mathbf{M}_\ell \frac{\hat{\partial}}{\hat{\partial}x_\ell} \frac{\hat{\partial}}{\hat{\partial}t} \mathbf{q}_j + \sum_\ell \mathbf{M}_\ell \frac{\hat{\partial}}{\hat{\partial}x_\ell} \sum_r \frac{\hat{\partial}}{\hat{\partial}x_r} \mathbf{F}_r(\mathbf{q}_j) = \mathbf{0}.$$

By applying the identities (45) and (46), we find for the double summation term:

$$\begin{aligned}
 \sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \sum_r \frac{\hat{\partial}}{\partial \hat{x}_r} \mathbf{F}_r(\mathbf{q}_j) &= \sum_{\ell, r} \frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \frac{\hat{\partial}}{\partial \hat{x}_r} \mathbf{M}_{\ell} \mathbf{F}_r(\mathbf{q}_j) + O(\Delta x^p) \\
 (53) \qquad \qquad \qquad &= \sum_{\ell, r} \frac{\hat{\partial}}{\partial \hat{x}_r} \frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \mathbf{M}_{\ell} \mathbf{F}_r(\mathbf{q}_j) + O(\Delta x^p) + O(\Delta x^{p-1})
 \end{aligned}$$

By using identities (53) and the antisymmetric condition (48), we get

$$(54) \qquad \qquad \qquad \sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \sum_r \frac{\hat{\partial}}{\partial \hat{x}_r} \mathbf{F}_r(\mathbf{q}_j) = O(\Delta x^{p-1}).$$

Therefore, by using the identities for the commutators (44) and (46), we can rewrite (52) as

$$\begin{aligned}
 \mathbf{0} &= \sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \frac{\hat{\partial}}{\partial t} \mathbf{q}_j + O(\Delta x^{p-1}) \\
 (55) \qquad \qquad \qquad &= \frac{\hat{\partial}}{\partial t} \sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \mathbf{q}_j + O\left(\frac{\Delta x^p}{\Delta t}\right) + O\left(\frac{\Delta t^q}{\Delta x}\right) + O(\Delta x^{p-1}),
 \end{aligned}$$

which is equivalent to (49). \square

This theory is valid in the case of sufficiently smooth solutions. A numerical scheme cannot distinguish between discontinuous solutions and smooth solutions with high gradients. So at the first glance, the Theorem directly transfers to that case. But due to stability reasons, one has to take measures to prevent unphysical oscillations, which results in the need of some limiting technique, like TVD, ENO/WENO etc. As a consequence of the application of limiting, the order of the scheme near discontinuities is lowered. Thus, the estimate (49) is much weaker near discontinuities than in smooth regions.

3.2.2. *The semi-discrete case.* For the semi-discrete case, we have to consider the construction via numerical flux functions in more detail. In the context of finite volumes, numerical schemes usually are represented in the semi discrete form

$$(56) \qquad \qquad \qquad \frac{\partial}{\partial t} \mathbf{q}_j - \sum_{i \in \mathfrak{R}_j} G_i(\mathbf{q}) = \mathbf{0},$$

where \mathfrak{R}_j denotes the set of all cell faces of cell \mathbf{j} , and g denotes a numerical flux function, normal to the cell face. This numerical flux function is allowed to depend on several \mathbf{q}_i , i. e. on the values of \mathbf{q} in several cells of the computational grid. Therefore, (56) represents a system of ordinary differential equations in time. When we solve this, using some standard scheme for ODE's, at a first glance the discrete time derivative only depends on values in the same space point. But this is not true for many choices of the numerical flux function g .

We now take a closer look at a typical example: One of the most important nonlinear schemes is the scheme by Harten, Lax, and van Leer [19], which for our purposes, is a nice model since it clearly distinguishes between the central and the upwinding part of the viscous flux. For this, the numerical flux function reads

$$(57) \quad G_{\text{HLL}}(\mathbf{q}_r, \mathbf{q}_\ell) = \frac{1}{2}(f(\mathbf{q}_r) + f(\mathbf{q}_\ell)) - \frac{1}{2} \frac{S_R + S_L}{S_R - S_L} (f(\mathbf{q}_r) - f(\mathbf{q}_\ell)) + \frac{S_R S_L}{S_R - S_L} (\mathbf{q}_r - \mathbf{q}_\ell)$$

with some bounding signal speeds S_L and S_R for the Riemann problem defined by the states left and right of the cell face, \mathbf{q}_ℓ and \mathbf{q}_r . If $S_R = -S_L = \frac{\Delta x}{\Delta t}$ for Cartesian equidistant

grids, this is just the numerical flux of the Lax-Friedrichs scheme. If we have a tighter but still symmetric choice $S_R = -S_L$ of the bounding speeds we find the Rusanov- or Local-Lax-Friedrichs scheme. In (57) there are three contributions: a symmetric one, that would leave us with central differences of second order in space, an upwinding term, and another symmetric term, that does not depend on the flux, but only on the state \mathbf{q} itself. The second and third terms both contribute to the numerical viscosity. If we write the resulting scheme in the fully discrete difference form

$$\frac{\hat{\partial}}{\hat{\partial}t} \mathbf{q}_j + \hat{\nabla} \cdot \mathbf{F}(\mathbf{q}_j) = \mathbf{0} ,$$

the third term becomes a part of the time difference instead of the space difference. For the semi-discrete scheme, the central viscosity terms make up an additional sum:

$$(58) \quad \frac{\partial}{\partial t} \mathbf{q}_j + \sum_r \frac{\hat{\partial}}{\partial x_r} F_r(\mathbf{q}_j) + \sum_{\mathbf{k} \in \mathbf{K}} \gamma_{\mathbf{k},j} \mathbf{q}_k = \mathbf{0} .$$

Thus, an analogue of Theorem 2 is true with (49) is replaced by

$$(59) \quad \frac{\partial}{\partial t} \left(\sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\partial x_{\ell}} \mathbf{q}_j \right) + \sum_{\mathbf{k} \in \mathbf{K}} \gamma_{\mathbf{k},j} \left(\sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\partial x_{\ell}} \mathbf{q}_k \right) = \mathcal{O}\left(\frac{\Delta x^p}{\Delta t}\right) + \mathcal{O}\left(\frac{\Delta t^q}{\Delta x}\right) + \mathcal{O}(\Delta x^{p-1}) ,$$

which can be interpreted as a discrete heat equation with a source term of order $\mathcal{O}\left(\frac{\Delta x^p}{\Delta t}\right) + \mathcal{O}\left(\frac{\Delta t^q}{\Delta x}\right) + \mathcal{O}(\Delta x^{p-1})$. With a suitable choice of the central part of the numerical viscosity, we can expect the discrete involution to converge to zero in time. With a poor choice, it might increase in time, even if the right hand side of (59) vanishes.

3.2.3. *A linear special case.* In this section, we consider a linear special case, which allows for exact discrete involutions. As a consequence of the previous sections, the approximation error in discrete involutions is mainly due to the commutators of the discrete differential operators. A smaller contribution is due do the commutator of these operators with the matrices \mathbf{M}_i , which make up the involution (2). If the commutators vanish, the involution is exact. We take a closer look at discrete differential operators that can be rewritten as

$$(60) \quad \frac{\hat{\partial}}{\hat{\partial}t} \mathbf{h}_j = \sum_{\mathbf{i} \in \mathbf{I}} \tilde{\alpha}_i \mathbf{h}_{j+\mathbf{i}}$$

for the time derivative and

$$(61) \quad \frac{\hat{\partial}}{\partial x_{\ell}} \mathbf{h}_j = \sum_{\mathbf{i} \in \mathbf{I}} \tilde{\beta}_i^{\ell} \mathbf{h}_{j+\mathbf{i}} ,$$

where \mathbf{i} and \mathbf{j} are index vectors and \mathbf{I} is a set of index vectors. This is a typical situation on structured grids, staggered or collocated. Here in addition, we require the coefficients $\tilde{\alpha}_i$ and $\tilde{\beta}_i^k$ to depend only on their index \mathbf{i} . Thus, the resulting scheme for a hyperbolic conservation law is linear. If now the coefficients commute with each other, for the mixed derivatives, which are just double summations, we find

$$(62) \quad \frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \left(\frac{\hat{\partial}}{\hat{\partial}t} \mathbf{h}_j \right) = \frac{\hat{\partial}}{\hat{\partial}t} \left(\frac{\hat{\partial}}{\partial \hat{x}_{\ell}} \mathbf{h}_j \right) .$$

They commute; the commutator vanishes. For the coefficients to commute with the \mathbf{M}_i we have the additional requirement that the \mathbf{M}_i are square matrices or the coefficients are scalar. So, in most cases we are restricted to scalar coefficients, especially for divergence

preserving transport. With these preparations we can state the following discrete analogue of 1:

Theorem 3. *Let the following (weakly) hyperbolic conservation law be given*

$$(63) \quad \mathbf{q}_t + \nabla \cdot \mathbf{F}(\mathbf{q}) = 0$$

together with constant matrices \mathbf{M}_i that satisfy

$$(64) \quad \mathbf{M}_\ell F_r + \mathbf{M}_r F_\ell = 0 \quad \forall \ell, r = 1, 2, \dots$$

Furthermore let the following linear difference operators be given

$$(65) \quad \frac{\hat{\partial}}{\hat{\partial} x_\ell} \mathbf{q}_j = \sum_{i \in I^\ell} \tilde{\beta}_i^\ell \mathbf{q}_{j+i},$$

$$(66) \quad \frac{\hat{\partial}}{\hat{\partial} t} \mathbf{q}_j = \sum_{\mathbf{k} \in \mathbf{K}} \tilde{\alpha}_{\mathbf{k}} \mathbf{q}_{j+\mathbf{k}},$$

and the coefficients $\beta_{\mathbf{k}}$ and α_i^ℓ commute with each other and with the \mathbf{M}_i .

If we discretize the conservation law (63) with the finite difference operators (65) and (66), then the following analogue of (4) holds true:

$$(67) \quad \frac{\hat{\partial}}{\hat{\partial} t} \left(\sum_{\ell} \mathbf{M}_\ell \frac{\hat{\partial}}{\hat{\partial} x_\ell} \mathbf{q}_j \right) = 0 \quad \forall \mathbf{j},$$

the discrete involution is exact.

If we assume the scheme to be constructed by means of numerical flux functions and consider the semi-discrete form, equation (67) has to be replaced by

$$(68) \quad \frac{\partial}{\partial t} \left(\sum_{\ell} \mathbf{M}_\ell \frac{\partial}{\partial x_\ell} \mathbf{q}_j \right) + \sum_{\mathbf{k} \in \mathbf{K}} \tilde{\gamma}_{\mathbf{k}} \left(\sum_{\ell} \mathbf{M}_\ell \frac{\partial}{\partial x_\ell} \mathbf{q}_{j+\mathbf{k}} \right) = \mathbf{0}.$$

This is true if the coefficients $\tilde{\beta}_i^\ell$ arising from the upwind part of the numerical flux function satisfy the above mentioned requirements: they commute with each other and with the \mathbf{M}_i . If, for example, the HLL-flux (57) is applied to a constant coefficient hyperbolic system, the resulting coefficients $\tilde{\beta}_i^\ell$ are scalar constants.

3.3. Discrete involutions and standard schemes. In the beginning of Sect. 3, we raised the question, which type of discrete involutions, if at all, we will find with standard schemes. Will we find exact ones or only approximate ones? Here we restrict our study to constant coefficient schemes. Thus, the only remaining question is: do the coefficients commute. We consider the Lax-Friedrichs, Lax-Wendroff, upwind-, and leapfrog scheme followed by a remark on the use of Runge-Kutta schemes for the time-discretization. If these schemes are applied to a constant coefficient hyperbolic system, which means $F_i(\mathbf{q}) = \mathbf{A}_i \mathbf{q}$, due to the constant signal speeds, a constant time-step can be chosen, so that not only the coefficients of the space discretization are constant, but also those of the time difference.

Although the schemes investigated in this section are rarely used in their pure form, most schemes in practical use are generalizations of these simple methods and, thus, inherit some of the properties of the underlying linear scheme. The results will be explored in Sect. 4 to study the interplay of discrete involutions and resonance.

3.3.1. *The Lax-Friedrichs scheme.* The behaviour of the Lax-Friedrichs scheme is best understood, if we take a careful look on its derivation. The starting point is the desire for a simple symmetric scheme. Therefore, the most obvious choice is to take central differences of second order in space and forward differences in time. In one space dimension, this leads to the simple explicit scheme

$$(69) \quad \frac{\mathbf{q}_k^{n+1} - \mathbf{q}_k^n}{\Delta t} + \frac{F(\mathbf{q}_{k+1}^n) - F(\mathbf{q}_{k-1}^n)}{2\Delta x} = 0 .$$

Since this turns out to be unconditionally unstable, one looks for a replacement. In the Lax-Friedrichs scheme this modification is done in a symmetric way. In the time discretization, the value \mathbf{q}_k^n is replaced by the arithmetic mean of its neighbours in space:

$$(70) \quad \frac{\mathbf{q}_k^{n+1} - \frac{1}{2}(\mathbf{q}_{k+1}^n + \mathbf{q}_{k-1}^n)}{\Delta t} + \frac{F(\mathbf{q}_{k+1}^n) - F(\mathbf{q}_{k-1}^n)}{2\Delta x} = 0 .$$

An interesting consequence of this construction is that \mathbf{q}_k^{n+1} does not depend on \mathbf{q}_k^n . An advantage of this is the possibility to use the scheme in a staggered manner, meaning that in each time step we toggle between evaluating at odd and even indexes. This gave rise to the development of the Nessyahu-Tadmor scheme [32]. As a disadvantage, in non staggered use of the scheme, high frequency oscillations are observed [7].

It is possible to rewrite the scheme in the usual conservation form, making the difference between formulas (69) and (70) part of the numerical flux function. In the sense of applying discrete difference operators instead of the analytic ones to the conservation law (1), this would result in an additional, viscous flux. But the difference, although made part of the numerical flux, still remains part of the time discretization because the correction term does not include any contributions of the flux function $f(\cdot)$. Thus, we have

$$(71) \quad \begin{aligned} \frac{\hat{\partial}}{\hat{\partial} t} \mathbf{q}_k &= \frac{\mathbf{q}_k^{n+1} - \frac{1}{2}(\mathbf{q}_{k+1}^n + \mathbf{q}_{k-1}^n)}{\Delta t} \\ &= \frac{\mathbf{q}_k^{n+1} - \mathbf{q}_k^n}{\Delta t} - \frac{\Delta x^2}{2\Delta t} \frac{\mathbf{q}_{k+1}^n + 2\mathbf{q}_k^n + \mathbf{q}_{k-1}^n}{\Delta x^2} . \end{aligned}$$

If we apply this discrete time derivative to a scalar quantity h , the condition

$$(72) \quad \frac{\hat{\partial}}{\hat{\partial} t} h_k = 0 \quad \forall k$$

is the same as applying a simple explicit method to the heat equation

$$(73) \quad h_t - \frac{\Delta x^2}{2\Delta t} h_{xx} = 0 .$$

If we solve heat equation (73) exactly, employing homogeneous Dirichlet conditions on the boundaries, we find that h converges to zero at any place. If, instead of the scalar h , we apply equation (72) to a vector quantity \mathbf{h} the same holds true for every component of \mathbf{h} . In several space dimensions, we get a spatial anisotropic heat equation, e. g.

$$\mathbf{h}_t - \frac{\Delta x^2}{2\Delta t} \mathbf{h}_{xx} - \frac{\Delta y^2}{2\Delta t} \mathbf{h}_{yy} - \frac{\Delta z^2}{2\Delta t} \mathbf{h}_{zz} = \mathbf{0}$$

for three dimensions. As a consequence, e. g. in the case of homogeneous Dirichlet boundary conditions, all components of \mathbf{h} converge to zero. Therefore, if we have a conservation law

$$\mathbf{q}_t + \nabla \cdot \mathbf{F}(\mathbf{q}) = 0$$

with an involution

$$\sum_i \mathbf{M}_i \mathbf{q}_{x_i} = 0 ,$$

discretized with the Lax-Friedrichs scheme and boundary conditions, which are consistent with the involution, then

$$\sum_i \mathbf{M}_i \frac{\hat{\partial}}{\partial x_i} \mathbf{q}$$

is an exact discrete involution, which even converges to zero in time.

Since the Balbas-Tadmor scheme [2] by its construction is close to the Lax-Friedrichs scheme, we can already at this point expect that it produces only small divergence errors, which are even nicely damped away.

We will use this considerations later on to identify in numerical flux functions the terms which have to be considered a contribution to the discrete time derivative instead of the space derivative. And we will employ a systematic control on these terms, namely the central viscosity, to minimize the production of divergence errors in a standard scheme.

3.3.2. *The Lax-Wendroff scheme.* To study the Lax-Wendroff scheme, we start with the simplest possible system of conservation laws: the scalar linear advection equation

$$(74) \quad q_t + a q_x = 0 .$$

The idea for the Lax-Wendroff scheme and its relatives is to start with a Taylor expansion in time:

$$(75) \quad q(x, t + \Delta t) = q(x, t) + \Delta t q_t(x, t) + \frac{1}{2} \Delta t^2 q_{tt}(x, t) + \mathcal{O}(\Delta t^3) .$$

Using the original conservation law (74) and its time derivative, the time derivatives in (75) can be replaced by space derivatives:

$$(76) \quad q(x, t + \Delta t) = q(x, t) - a \Delta t q_x(x, t) + \frac{1}{2} a^2 \Delta t^2 q_{xx}(x, t) + \mathcal{O}(\Delta t^3) .$$

From this we get the Lax-Wendroff scheme by applying standard second order central differences for first and second space derivatives. If we use standard upwind differences of second order, we find the Beam-Warming scheme. The arithmetic mean of both schemes results in the Fromm scheme.

Let us now concentrate on the Lax-Wendroff scheme. Since, according to the above choice, we have

$$(77) \quad \frac{\hat{\partial}}{\partial x} q_j = \frac{q_{j+1} - q_{j-1}}{2\Delta x} ,$$

for the discrete second space derivative, we would expect

$$(78) \quad \frac{\hat{\partial}}{\partial x} \left(\frac{\hat{\partial}}{\partial x} q_j \right) = \frac{q_{j+2} - 2q_j + q_{j-2}}{4\Delta x^2} .$$

But the Lax-Wendroff scheme employs

$$(79) \quad \frac{\tilde{\partial}}{\partial \tilde{x}^2} q_j = \frac{q_{j+1} - 2q_j + q_{j-1}}{\Delta x^2} ,$$

which is apparently not the same. To interpret this as a part of the space discretization, we would have to write it in terms of the difference operator (77). But this is impossible. Therefore, it is impossible to interpret the viscosity term of the Lax-Wendroff scheme as a part of the space discretization, even in the simple case of the one dimensional scalar

advection equation. Instead, we have to view it as a part of the time difference. Thus, the time difference would read as

$$\begin{aligned} \frac{\hat{\partial}}{\hat{\partial}t} q_j^n &= \frac{q_j^{n+1} - q_j^n}{\Delta t} + \frac{1}{2} a^2 \Delta t^2 \frac{q_{j+1}^n - 2q_j^n + q_{j-1}^n}{\Delta x^2} \\ &= \frac{1}{\Delta t} q_j^{n+1} - \left(\frac{1}{\Delta t} + a^2 \frac{\Delta t^2}{\Delta x^2} \right) q_j^n + \frac{1}{2} a^2 \frac{\Delta t^2}{\Delta x^2} (q_{j+1}^n + q_{j-1}^n). \end{aligned}$$

A similar formula would be found for a one-dimensional linear system of conservation laws. But then, we would have to replace a by the system matrix \mathbf{A} . Thus, the coefficients in the discrete time derivative become matrix valued. So, Theorem 3 can only be applied to a small number of systems, namely those, for which the system matrix \mathbf{A} and the matrix \mathbf{M} which makes up the involution commute.

If we had used (78) instead of (79) for the second derivative, it would have been possible to interpret the viscous term as a part of the discrete space derivative. But in the case of a system this, again, leads to matrix valued coefficients – this time in the discrete space derivative. Thus, the same restrictions apply as for the original Lax-Wendroff scheme. In addition, for systems in several space directions we would have to require the matrices \mathbf{A}_i for the different space directions to commute with each other.

For systems in several space dimensions we only show a two-dimensional example

$$(80) \quad \mathbf{q}_t + \mathbf{A}\mathbf{q}_x + \mathbf{B}\mathbf{q}_y = \mathbf{0}.$$

For this the analogue of (76) reads as

$$(81) \quad \begin{aligned} \mathbf{q}(x, y, t + \Delta t) &= \mathbf{q}(x, y, t) - \Delta t(\mathbf{A}\mathbf{q}_x + \mathbf{B}\mathbf{q}_y) \\ &+ \frac{1}{2} \Delta t^2 (\mathbf{A}^2 \mathbf{q}_{xx} + \mathbf{A}\mathbf{B}\mathbf{q}_{yx} + \mathbf{B}\mathbf{A}\mathbf{q}_{xy} + \mathbf{B}^2 \mathbf{q}_{yy}) + O(\Delta t^3). \end{aligned}$$

Apparently, the same arguments hold as for one space dimension. If we take the viscous term as part of the time difference, we can apply Theorem 3, as long as both of \mathbf{A} and \mathbf{B} commute with both of the matrices \mathbf{M}_x and \mathbf{M}_y making up an involution of system (80). This extends to higher dimensions in a straight forward manner.

If the matrices do not commute, we only find – provided the viscous term is taken as part of the time difference –

$$(82) \quad \sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\hat{\partial}t} \frac{\hat{\partial}}{\hat{\partial}x_{\ell}} \mathbf{q}_j = \mathbf{0}.$$

This is a much weaker condition than (67). In fact, numerical experiments show that the approximation of the involution is in no way better than for any nonlinear scheme of the same order.

For the Beam-Warming scheme, the results are quite similar. Now, most second order schemes, especially those based on TVD-limiters, are constructed by using weighted means of the Lax-Wendroff, as a central scheme, and the Beam-Warming, as an upwind scheme. Thus, for these schemes, we can not expect the conditions of Theorem 3 to hold. The best we can hope for, is an approximate involution in the sense of Theorems 2 and 1.

3.3.3. The upwind-scheme. For a scalar conservation law, the upwind scheme assigns a one sided difference operator to each space derivative. This operator takes into account the upwind direction, i. e. for positive signal speed, backward differences are used and for negative signal speeds forward differences. In the case of a linear system, the upwind method is applied to each characteristic field.

The simple case: full upwinding. The simplest case is full upwinding: in each space direction for all characteristic fields the same upwind direction is found. In this case all discrete space derivatives $\hat{\partial}/\hat{\partial}x_r$ are one sided standard differences of first order, forward or backward, depending on the upwind direction for x_r .

The effects of this can be nicely seen, when the scheme is applied to the linearized induction equation of two-dimensional magnetohydrodynamics:

$$(83) \quad \mathbf{B}_t - \nabla \times (\mathbf{v} \times \mathbf{B}) = 0, \quad \mathbf{v} = (u, v)^T \equiv \text{constant}$$

with positive velocity components u and v . As Fey and Torrilhon [39] point out, this is an interesting example, modeling most of the important properties of real MHD, at least in the context of involutions. It is a linear conservation system with $\nabla \cdot \mathbf{B}$ as an involution. With the matrices $\mathbf{M}_1 = (1 \ 0)$ and $\mathbf{M}_2 = (0 \ 1)$, we find that it satisfies the conditions for Theorem 1. Thus, with appropriate difference operators, we will obtain a discrete involution.

In space, we employ two different types of differences. First we use standard upwind. Since there is only one nonzero wave speed for each space direction, we end up just with one-sided differences for $\hat{\partial}/\hat{\partial}x$ and $\hat{\partial}/\hat{\partial}y$. So we have no matrix valued coefficients, and the conditions of Theorem 3 are satisfied. For a second test, we employ the Corner-Transport-Upwind (CTU) scheme, a variant of standard upwind, which takes into account the direction of the transport. This results in the transverse upwind differences

$$(84) \quad \begin{aligned} \frac{\hat{\partial}}{\hat{\partial}x} h &= (1 - c_y) \frac{h_{i,j} - h_{i-1,j}}{\Delta x} + c_y \frac{h_{i,j-1} - h_{i-1,j-1}}{\Delta x}, \\ \frac{\hat{\partial}}{\hat{\partial}y} h &= (1 - c_x) \frac{h_{i,j} - h_{i,j-1}}{\Delta y} + c_x \frac{h_{i-1,j} - h_{i-1,j-1}}{\Delta y}, \end{aligned}$$

where c_x and c_y denote the directional Courant numbers. In time, we always employ forward differences of first order. Therefore we expect the involution to be constant in time.

First example. As initial data, we discretize the divergence free field $\mathbf{B} = (B_1, B_2)^T$ with

$$B_1 = \cos(2\pi x + \pi y) \quad B_2 = -2 \cos(2\pi x + \pi y)$$

on a 320×320 grid for the square region $[-1, 1] \times [-1, 1]$ with periodic boundary conditions. For the discrete initial values, we employ a rather naive method: we just evaluate at the cell center. Thus, the initial divergence is not exactly zero. The results are shown in Figure 1. In the left picture we see that the discrete divergence measured in upwind differences is constant in time, it sticks to its initial value, if the standard upwind scheme is used. In the right picture, the same is found for the divergence measured in transverse differences with the according CTU-scheme employed. Although not depicted here, in both cases not only the norm of the divergence is constant. The discrete divergence itself is constant, as was predicted by the above theory.

The divergence measured in central differences, although almost zero in the initial state, grows to approach the divergence measured in terms of the difference operator used in the scheme, which is indeed much larger.

From this, we can draw two important conclusions: First, the usual technique of projecting the magnetic field to a divergence free field with respect to some higher order central difference is insufficient. The projection should be done with respect to the difference operator actually used in the scheme. For general nonlinear systems with changing upwind directions, this is nearby impossible. Especially, it is impossible to provide a ‘‘divergence free’’ initial state that is adequate for all cases. Second, upwind schemes, by their lack of central viscosity, are unable to damp the divergence error introduced by the initial state.

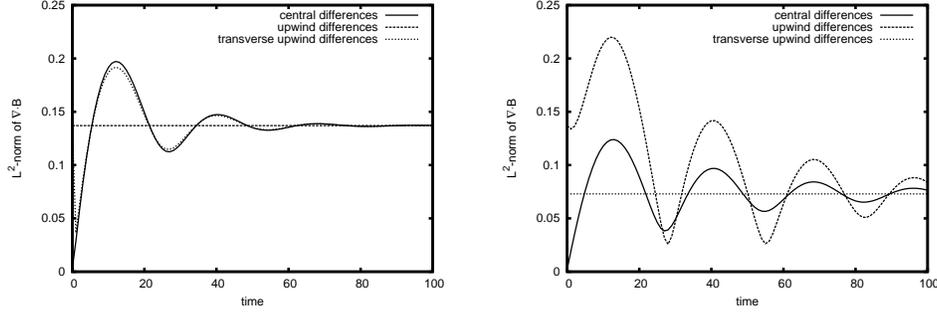


FIGURE 1. Smooth example: time behaviour of the L^2 -norm of the discrete divergence for standard upwind (left) and corner transport upwind (right), measured with central differences, upwind differences, and transverse upwind differences.

Second example. As a second example, we present an oblique Riemann-problem, a piecewise constant initial state with discontinuity normal to $(1, 1)^T$ reproduced on a Cartesian grid. The discontinuity is just the diagonal of the cells it intersects. For the left and right state and the state in the cells with the discontinuity, we take

$$(85) \quad \mathbf{B}_l = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbf{B}_r = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \text{and thus} \quad \mathbf{B}_r^* = \mathbf{B}_l^* = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$$

for the cells intersected by the discontinuity, i. e. we project the data onto the grid in a finite volume manner. The data are analytically divergence free. For $u, v > 0$ they are also discrete divergence free when we employ upwind differences. For $u > 0, v < 0$, they are not.

In Figure 2 it can be seen that also in the discontinuous case the divergence measured in the differences used in the scheme is constant. Figure 2 also shows that the initial state has to be divergence free with respect to the differences used in the scheme. If not, the divergence will raise pretty soon. The worst results are obtained, when we do a wrong upwinding (lower row). For linear systems like our model problem, this is no issue. But for nonlinear systems like full MHD, this adds a new problem to the lack of exact involutions: Since the upwind direction depends on the state, it is in general impossible to know the difference operators in advance. So, the best we can get is an initial divergence in the order of the scheme itself.

The general case. For the investigation of the general case, we start with a one dimensional situation:

$$(86) \quad \mathbf{q}_t + \mathbf{A}\mathbf{q}_x = 0.$$

For a hyperbolic conservation law, \mathbf{A} can be decomposed into

$$(87) \quad \mathbf{A} = \mathbf{R}\mathbf{\Lambda}\mathbf{L},$$

where \mathbf{R} and $\mathbf{L} = \mathbf{R}^{-1}$ are the matrices of the right and left eigenvectors of \mathbf{A} and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues of \mathbf{A} . By manipulating the entries of $\mathbf{\Lambda}$ one can easily construct matrices \mathbf{A}^+ , \mathbf{A}^- and $|\mathbf{A}|$ which have the same eigenvectors as \mathbf{A} but differ in their eigenvalues: For \mathbf{A}^+ all negative eigenvalues are replaced by zero, for \mathbf{A}^- the positive ones,

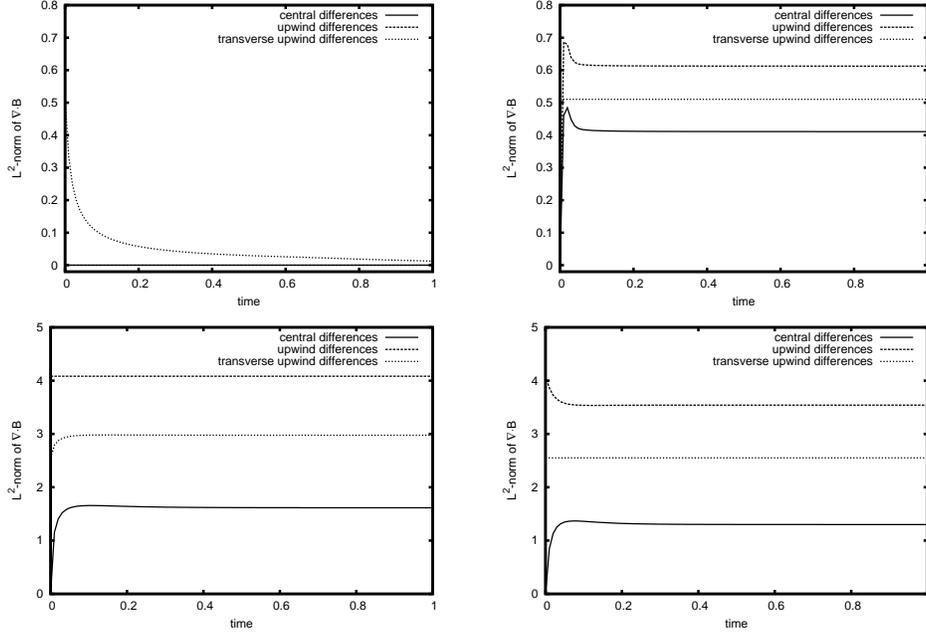


FIGURE 2. Discontinuous example: time behaviour of the L^2 -norm of the discrete divergence for standard upwind (left) and corner transport upwind (right), measured with central differences, upwind differences, and transverse upwind differences. Upper row: original setting; lower row: sign of v changed

and for $|\mathbf{A}|$ we replace all eigenvalues by their absolute values. Using these matrices, we can write the resulting discrete space difference operator as

$$\begin{aligned}
 (88) \quad \frac{\hat{\partial}}{\partial x}(\mathbf{A}\mathbf{q}_k) &= \mathbf{A}^+ \frac{\mathbf{q}_k - \mathbf{q}_{k-1}}{\Delta x} + \mathbf{A}^- \frac{\mathbf{q}_{k+1} - \mathbf{q}_k}{\Delta x} \\
 &= \frac{1}{\Delta x} [\mathbf{A}^+ \mathbf{q}_k - \mathbf{A}^+ \mathbf{q}_{k-1} + \mathbf{A}^- \mathbf{q}_{k+1} - \mathbf{A}^- \mathbf{q}_k] \\
 &= \frac{1}{\Delta x} [-\mathbf{A}^+ \mathbf{q}_{k-1} + |\mathbf{A}| \mathbf{q}_k + \mathbf{A}^- \mathbf{q}_{k+1}] \\
 &= -\frac{1}{2\Delta x} [(|\mathbf{A}| + \mathbf{A}) \mathbf{q}_{k-1} - 2|\mathbf{A}| \mathbf{q}_k + (|\mathbf{A}| - \mathbf{A}) \mathbf{q}_{k+1}].
 \end{aligned}$$

From these manipulations it can be easily seen that it is impossible to write the difference operator without matrix valued coefficients. Therefore, Theorem 3 can only be applied if \mathbf{A} and the matrix \mathbf{M} for the involution commute. In the multidimensional case, we have to require that all \mathbf{A}_r and \mathbf{M}_ℓ commute with each other. Thus, in general, we find no exact involutions for the upwind scheme, especially when the involution is a divergence.

Since most high quality numerical flux functions are based on upwinding, this implies that in real world computations, we can only expect an approximate involution in the sense of Theorem 2. In addition, the lack of central viscosity prevents the scheme from damping the errors in the involution.

3.3.4. *A remark on the use of Runge-Kutta schemes.* Runge-Kutta schemes play an important role in numerical simulations of time-dependent problems. They are also the method of choice for the starting procedure in a multistep scheme like leapfrog and its variants. Therefore, we are interested in the effects of using them for systems with involutions.

If the space discretization is done with differences satisfying the conditions of Theorem 3, then we get

$$(89) \quad \frac{\partial}{\partial t} \left(\sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\partial x_{\ell}} \mathbf{q}_{\mathbf{j}} \right) = \mathbf{0} .$$

If a consistent one step method is applied to that, the resulting scheme is involution preserving. When taken as a starting procedure for leapfrog, it also leads to an involution preserving scheme.

If the scheme is constructed by means of numeric flux functions, we get for the semi-discrete involution the expression given in (68) if the requirements given there are satisfied. This expression includes the central numerical viscosity. It corresponds to a discretized parabolic equation. When the numerical viscosity is reasonable, any stable time discretization shows the same behaviour as we found in Sect. 3.3.1 for the involution in the Lax-Friedrichs scheme.

4. DISCRETE INVOLUTIONS AND RESONANCE

In this section we identify discrete involutions and resonance as they key one needs to understand how divergence errors arise in MHD-simulations and destroy them.

By means of a computational example we show how resonance makes the estimates for the discrete involution in Theorem 2 worthless. We study the role of the central viscosity of the scheme and explain why the Balbas-Tadmor scheme [2, 1] and the Zachary-Malagoli-Colella scheme [41] produce only small divergence errors. In this course, we present a modification of the Roe-solver which shows the same stability. This modification is not intended to replace divergence cleaning, but to reduce the errors which have to be swept out of the computational domain.

4.1. **The de Sterck test.** The De Sterck test [36] is a special configuration for a shallow water MHD flow. It shows a strong tendency to develop resonant phenomena and, thus, to single out numerical schemes which are prone to divergence errors. The test problem imposes a supersonic horizontal grid-aligned inflow on the left boundary of a rectangular domain. The initial state in the lower half of the domain, and also of the left boundary, contains a resonant mode. The initial data in the upper half are

$$(90) \quad h = 2 , \quad u = 5.5 , \quad v = 0 , \quad B_1 = 0.5 , \quad B_2 = 0 ,$$

and in the lower half

$$(91) \quad h = 1 , \quad u = 4.5 , \quad v = 0 , \quad B_1 = 2 , \quad B_2 = 0 .$$

The gravitational constant is set to one. Since the discontinuity is aligned with the grid, the initial data are discrete divergence free for any reasonable difference operator. We performed a test on a 200×200 grid for the domain $[-1, 1] \times [-1, 1]$ with the Local Lax-Friedrichs scheme (LLF). The numerical flux over the cell faces is computed with 1d-physics. This is a widespread approach. In one-dimensional physics a one-dimensional divergence constraint applies. Thus the equation for hB_1 , in the full MHD the equation for B_1 , can be eliminated. The component hB_1 , or B_1 in full MHD, is constant in space and time and, thus, only a parameter. The reduced 1d-system is fully hyperbolic. When used

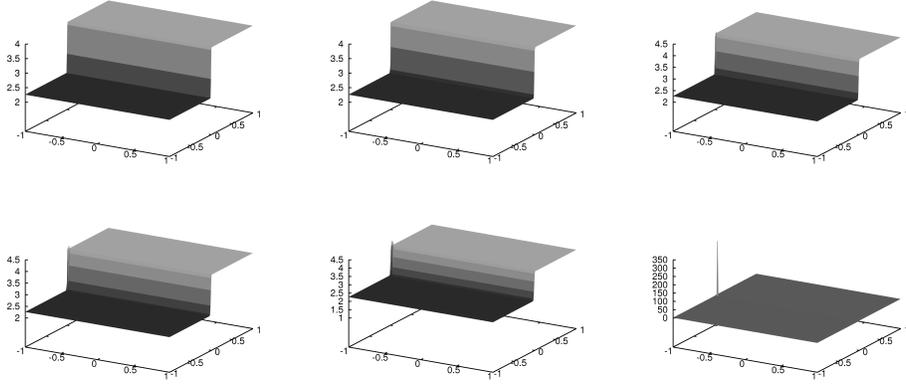


FIGURE 3. Absolute value of $u - c_g$ for De Sterck test with LLF based on one-dimensional physics. First six time steps (top row 1–3, bottom row 4–6). Note the different scaling in the last picture.

for multi-dimensional simulations, this introduces two difficulties: on each cell face the parameter for the magnetic field component normal to the face has to be chosen in some way, and we lose control over part of the viscosity of the scheme, namely the viscosity on the neglected wave. But this is exactly the wave which is responsible for resonance. For the first six time steps the absolute value of the resulting fastest wave speed, $u - c_g$, with the magnetogravitational speed $c_g = \sqrt{B_1^2 + gh}$ is plotted in Figure 3. It turns out that, in this case, resonance, once initiated, grows very fast. It also affects the wave speeds, which depend on the magnetic field. When we consider the estimate for the general discrete involution

$$(92) \quad \frac{\hat{\partial}}{\hat{\partial}t} \left(\sum_{\ell} \mathbf{M}_{\ell} \frac{\hat{\partial}}{\hat{\partial}x_{\ell}} \mathbf{q}_{\mathbf{j}} \right) = \mathcal{O}\left(\frac{\Delta x^p}{\Delta t}\right) + \mathcal{O}\left(\frac{\Delta t^q}{\Delta x}\right) + \mathcal{O}(\Delta x^{p-1}) \quad \forall \mathbf{j}$$

from Theorem 2, we find that the first order term, namely $\mathcal{O}(\Delta x^p/\Delta t)$ is most critical. The fast growing wave speeds result in a fast decreasing time step. Thus, the estimate (92) becomes weaker each time step. Divergence errors drive resonance, and resonance weakens the bound for the growth of the divergence errors.

In computations on Cartesian grids it is common to configure the initial state in a way that all discontinuities are aligned with the grid. For a piecewise constant initial state, consistent with the constraint, this means that for any consistent difference operators the discrete initial state also satisfies the discrete constraint. The involution can only be violated by rounding errors. Since rounding errors are $\mathcal{O}(1)$, the introduced error in the involution is of order $\mathcal{O}(1/\Delta x)$. Grid refinement results in even stronger resonance phenomena. The numerical viscosity and, thus, the damping of the resonance is reduced. Hence, for a scheme which fails due to resonance, it is impossible to improve the situation by grid refinement. The situation is even worse, as can be verified by the numerical tests in Sect. 4.4.

4.2. The role of the central numerical viscosity. To study the role of central numerical viscosity in more detail we begin with a simple example. In Figure 4, we show numerical results for the situation described in the beginning of Sect. 2.3.1. We trigger resonance by a jump of B_1 in the middle of the computational domain. Apparently the resonance effects

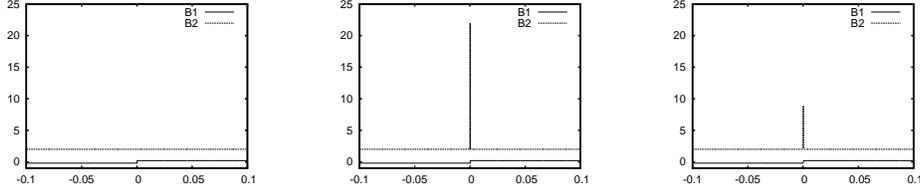


FIGURE 4. Effects of resonance: initial condition, result with CIR-scheme, result with Lax-Friedrichs scheme

are much weaker if we employ the Lax-Friedrichs scheme instead of the CIR scheme. The main difference between these two schemes is that the LF-scheme is central while the CIR scheme employs wave wise upwinding. Thus, the LF-scheme provides central viscosity, while the CIR-scheme does not.

But in the general case, the Lax-Friedrichs scheme, due to its high viscosity, is not preferable. Therefore, in practical use other schemes based on numerical flux functions are used. At this point, we reconsider the HLL-flux (57):

$$(93) \quad G_{\text{HLL}}(\mathbf{q}_r, \mathbf{q}_\ell) = \frac{1}{2}(f(\mathbf{q}_r) + f(\mathbf{q}_\ell)) - \frac{1}{2} \frac{S_R + S_L}{S_R - S_L} (f(\mathbf{q}_r) - f(\mathbf{q}_\ell)) + \frac{S_R S_L}{S_R - S_L} (\mathbf{q}_r - \mathbf{q}_\ell).$$

Obviously, the viscosity terms are closely related to the signal speeds. This is a general issue [20, 24, 12, 18]. Therefore, in practise, the central viscosity can not be chosen arbitrarily high. A simple approach is HLL with $S_L = -S_R$, which refers to the Local-Lax-Friedrichs scheme. This choice imposes a lower bound on the viscosity for all waves, thus also for the resonant wave². This is a prototype for many schemes, which do not explicitly resolve the resonant wave. Both, the Balbas-Tadmor scheme [2] and the Zachary-Malagoli-Colella scheme [41] belong to this class.

As a prototype of schemes which, by construction, explicitly resolve all waves, we consider the Harten entropy-fix [17] for the Roe-solver – not to be confused with the Harten-Hyman entropy-fix [18] –, which allows to impose a lower bound for the viscosity on each wave separately. It is constructed such that the viscosity depends smoothly on the wave speeds. Harten replaces the absolute value of an eigenvalue λ of the Roe-matrix by

$$(94) \quad \phi(\lambda) = \begin{cases} |\lambda| & \text{if } |\lambda| \geq \delta, \\ (\lambda^2 + \delta^2)/(2\delta) & \text{if } |\lambda| < \delta, \end{cases}$$

where δ is a small parameter. The numerical viscosity is bounded below by $\delta/2$. Since additional numerical viscosity on a single wave is equivalent to the splitting of the wave into two weaker waves [24, 18], the optimal, i. e. the maximal admissible, choice for the parameter is twice the largest absolute value of an eigenvalue of the Roe-matrix: $\delta = 2|\lambda_{\max}|$. This puts the same amount of viscosity on the wave as in the LLF scheme. A simpler, but still reasonable choice would be $\delta = 2|u|$. The speeds of the waves resulting from the corresponding splitting of the original resonant wave would be $\pm\lambda_{\max}$ or $\pm u$ respectively.

4.3. The assumption of one-dimensional physics in flux computations. To study the role of the assumption of one-dimensional physics in the construction of numerical flux functions, we start with an example. In section 4.1, we demonstrated the effects of resonance by applying the LLF-scheme with the numerical flux based on one-dimensional

²Since resonance only occurs in certain physical states, it would be more correct to call it the *wave which might become resonant*. But for the sake of readability, we stick to this simplistic formulation.

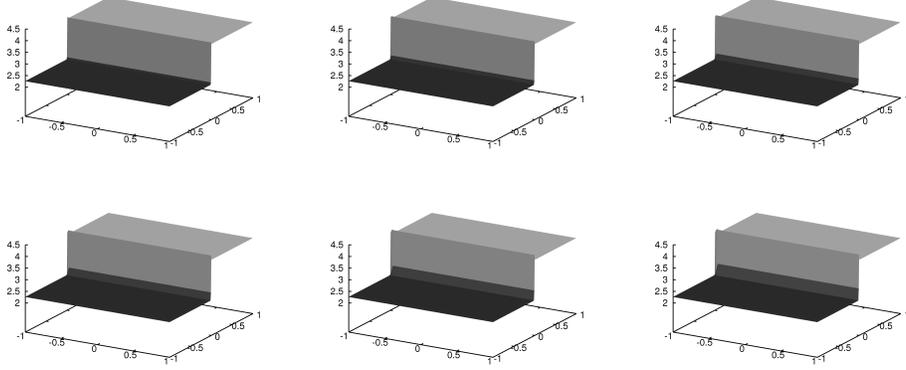


FIGURE 5. Absolute value of $u - c_g$ for De Sterck test with LLF for full system. First six time steps (top row 1–3, bottom row 4–6).

physics to the De Sterck test case. Now we repeat the same computation without the assumption of one-dimensional physics. The results are presented in Figure 5. As one would expect by the considerations of the previous section, the resonance is nicely damped. The divergence errors are much smaller than with the one-dimensional physics in Figure 3. Something got lost by the assumption of one-dimensional physics. The resulting viscosity seems to be weak or even antidiffusion on the resonant wave. Thus, in a scheme which uses projection to prevent divergence errors, the projection has to be done more often to keep the simulation stable. The work, saved by the easier flux computation, results in a much higher work for divergence cleaning.

The assumption of one-dimensional physics in the flux computation would, on a Cartesian grid, imply that all terms including B_{1x} , B_{2y} and B_{3z} are neglected. In general, this leads to a modeling error and, thus, to an error of order $\mathcal{O}(\Delta x^{-2})$ in numerical simulations. But in standard implementations of MHD it is still at least of order $\mathcal{O}(\Delta x)$. This can be verified by the following considerations:

We restrict our analysis to the x -direction in a Cartesian grid. In most codes the choice of the parameter B_1 is done in dependence on its values in the cells neighbouring to the cell face at which the flux has to be evaluated. Usually it is taken to be a weighted mean of these values. Thus, for the resulting full flux function we still have, if written for some one-dimensional situation, at the i -th interface,

$$(95) \quad G(\mathbf{q}_{i-l}, \dots, \mathbf{q}_{i+k}) \rightarrow F_1(\mathbf{q}), \quad \text{if } \mathbf{q}_{i+r} \rightarrow \mathbf{q} \quad \text{for } r = -l, \dots, k.$$

Hence, the flux function and, by applying the Lax-Wendroff theorem, the scheme itself is consistent. In smooth regions this implies an order of at least one. In addition, the error introduced to the antisymmetric condition (3) (when applied to \mathbf{G} instead of \mathbf{F}) is small. The actual order of such schemes can only be tested by measuring the experimental order of convergence (EOC). There is no direct control on the differences used. As a matter of experience, these schemes are most prone to failure due to divergence errors. The schemes by Zachary, Malagoli and Colella [41] and Balbás and Tadmor[2], mentioned in the introduction, do not employ the assumption of one-dimensional physics at any place.

For our prototype system, the linearized induction equation (20) in two space dimensions, the flux in x -direction is $(0, -vB_1 + uB_2)^T$. Let us assume that u is positive and we

employ the upwind scheme. The flux-term uB_2 is always treated with upwind differences. If we take the parameter B_1 to be the value in the cell left of the cell face for which the numerical flux is to be computed, we end up with full upwinding, and, according to Theorem 3, find an exact discrete involution. If we take the value of B_1 from the cell right of the cell face, the flux-term $-vB_1$ is discretized with downwind differences. The conditions of Theorem 3 are not longer valid. If we define $\hat{\partial}/\hat{\partial}x$ to be the upwind difference operator and $\tilde{\partial}/\tilde{\partial}x$ to be the downwind operator, the actual discretization for the second flux component at a fixed grid point x_i reads

$$\begin{aligned}
 (96) \quad \frac{\hat{\partial}}{\hat{\partial}x}(uB_2)_i - \frac{\tilde{\partial}}{\tilde{\partial}x}(vB_1)_i &= \frac{\hat{\partial}}{\hat{\partial}x}(-vB_1 + uB_2)_i - \left[\frac{\hat{\partial}}{\hat{\partial}x} - \frac{\tilde{\partial}}{\tilde{\partial}x} \right] (vB_1)_i \\
 &= \frac{\hat{\partial}}{\hat{\partial}x}(-vB_1 + uB_2)_i - v \frac{B_{1i+1} - 2B_{1i} + B_{1i-1}}{\Delta x} \\
 &= \frac{\hat{\partial}}{\hat{\partial}x}(-vB_1 + uB_2)_i - v\Delta x(B_{1,xxi} + O(\Delta x^2)).
 \end{aligned}$$

A similar consideration can be done for the y -direction. Summed up, the divergence error introduced in one time step is of order $O_s(\Delta x)$, which means that Δx is in turn of the same order as the divergence error. If instead of the value right to the cell face, we take a weighted mean with weight α for the right hand value, the error is just multiplied by α but still of the same order.

This is not too bad. Thus, the main reason for the problems arising from one-dimensional physics is the loss of control on the numerical viscosity on the resonant wave.

4.4. Numerical experiments. In this section, we present some numerical experiments³ for the De-Sterck test with a Roe-type scheme without the assumption of one-dimensional physics. Analytically, the problem results in a steady state, which has been already reached at time $t = 0.8$. To study the long-term effects, we went on to time $t = 4.8$. The left half of Figure 6 gives a comparison of the scheme with and without entropy fix. As entropy fix, we employ the above mentioned Harten fix with parameter $\delta = 2|\lambda_{\max}|$ or $\delta = 2|u|$ for the resonant wave and $\delta = 10^{-8}$ for the other waves. As Figure 6 shows, the effects of the central viscosity introduced by the entropy fix are strong. While the computation without the fix does not even reach the steady state, the computation with the fix survives the whole simulation without the need of an intermediate projection step. The choice $\delta = 2|u|$ is weaker, but still yields reasonable results. The right half of Figure 6 demonstrates the influence of the limiter on the stability. Although the limiter does not change anything on the resonant wave itself, since it propagates with zero speed, the choice of limiters for the other waves show some effect. For short times, the more compressive limiters, see [22], yield better results. But the unphysical forces arising from the divergence errors are much better resolved. The better resolution of discontinuities results in steeper gradients and, thus, in higher divergence errors. In the long-term run, the error exceeds the error obtained with the classical MC-Limiter.

Next, we investigate the influence of the grid resolution and the order of the scheme. On the one hand, a higher grid resolution and a higher order would, by the estimate (49) in Theorem 2, we would expect a positive effect. But on the other hand, the higher resolution and the higher order lower the numerical viscosity and allow for steeper gradients and, thus, for higher divergence errors near shocks. As Figure 7 shows, the second argument dominates for the grid resolution. On a fixed grid, the higher order scheme performs better.

³Numerical experiments in this paper are done with clawpack [8].

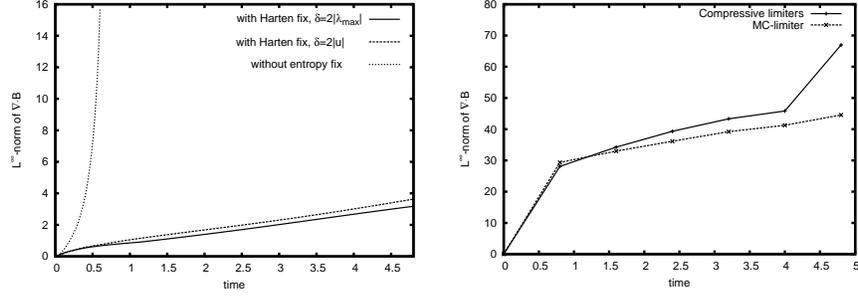


FIGURE 6. Left: Maximum norm of $\nabla \cdot (h\mathbf{B})$ over time for the De Sterck test problem with and without Harten fix for 1st order computation on 10×10 grid. Right: Maximum norm of $\nabla \cdot (h\mathbf{B})$ for 2nd order with Harten fix on 200×200 grid with standard MC-Limiter and highly compressive limiters.

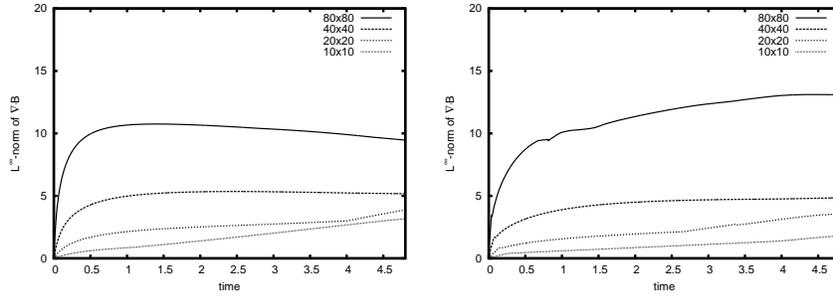


FIGURE 7. Influence of the grid resolution on the divergence error. Left: 1st order; Right: 2nd order.

Figure 8, finally, presents results of highly resolved computations, 200×200 grid cells, with the second order schemes. The basic structure of the solution is preserved even for the long-term run. But the divergence errors have infected all of the lower half of the computational domain. At the places with the highest divergence errors, disturbances of the solution can be seen in the contour plot of the height. The computations with the high resolving limiters in Figure 8 show an area with severe destruction of the solution. With the MC-limiter, this effect is weaker.

The situation is the same as for the schemes by Balbas and Tadmor [2, 1] and Zachary, Malagoli, and Colella [41]. It is still reasonable to employ some sort of divergence cleaning. But one can resort to a weaker one. In the case of a projection to a divergence free field, the time interval between two projections can be considerably increased, since the computation is still stable. In a scheme based on hyperbolic or mixed type GLM-divergence-cleaning [11], the divergence errors which have to be transported out of – and thus through a significant part of – the computational domain are much smaller.

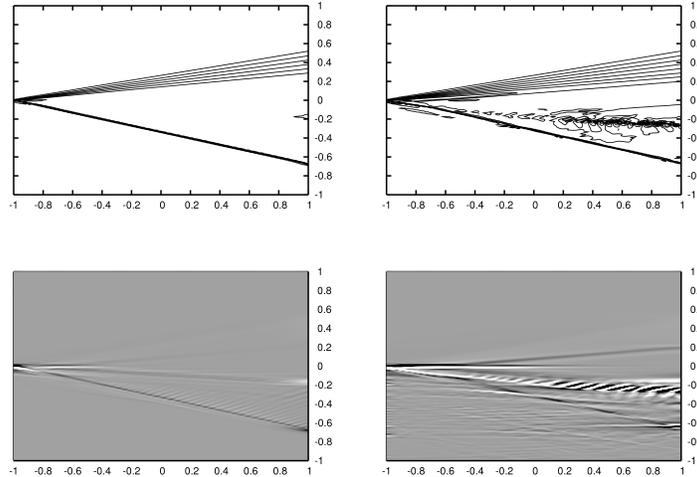


FIGURE 8. De-Sterck test with Harten fix on 200×200 grid for second order computation with highly compressive limiters at time $t = 0.8$ (left) and $t = 4.8$ right. Upper row: height; lower row: divergence

5. CONCLUSIONS AND OUTLOOK

In this study, we investigated the origin of divergence errors in MHD simulations. The concept of involutions, introduced by Dafermos [10, 9], turned out to be the key of understanding of the issue. Especially when, like in MHD, the involutions are closely related to resonance, their exact reproduction in the discrete case is needed to prevent the numerical schemes from failing due to unphysical forces. If an involution satisfies Dafermos' sufficient condition (3), discrete analogues of Theorem 1 give quantitative information on the possible errors. For some linear schemes, the discrete involutions are even exact. The introduction of central viscosity in the scheme provides a tool to reduce resonant effects. It turns the discrete involution into a parabolic equation, which damps the involution and, for example in the case of MHD, the resonance. But this only works if for the computation of the intercell fluxes the full multidimensional physics is taken into account. If the intercell fluxes are computed with the assumption of one-dimensional physics, in addition to not explicitly resolving the resonant wave, we completely neglect it. The resulting central viscosity cannot be controlled and, thus, be even of the wrong sign. There is simply no possibility to control it. Employing fluxes with full physics, like in the Balbas Tadmor scheme [1] and the Zacharay Malagoli Colella scheme [41], considerably stabilizes the scheme. In Roe-type schemes, we can explicitly tune the amount of central viscosity introduced by the flux function. If we employ the maximal admissible amount of viscosity on the resonant wave, the scheme is stable even for very long runs. Due to the disturbances of the solution, which are caused by the growing divergence errors, it is still reasonable to employ some sort of divergence cleaning. But one can resort to a weaker one. In the case of a projection to a divergence free field, the time interval between two projections

can be considerably increased. The computation is still stable. In a GLM-scheme [11], the disturbances introduced by the transport of divergence errors through the computational domain are minimized.

In summary, divergence errors in MHD are mainly caused by resonance and a lack of positive central viscosity in the applied numerical scheme, where the second most often results from the assumption of one-dimensional physics in the calculation of intercell fluxes.

Acknowledgements. First, I would like to thank Prof. G. Bader, whose probing questions gave rise to this study. Many thanks also to the participants of the 8th Hirschegg Workshop on Conservation Laws, especially to Manuel Torrilhon, for the deep inspiring discussions on the subject. I am also grateful to Felix Rieper for many interesting discussions and the proofreading of the manuscripts.

REFERENCES

1. Jorge Balbás, personal communication at the 12th Conference on Hyperbolic Problems, University of Maryland, June 2008.
2. Jorge Balbás and Eitan Tadmor, *Nonoscillatory central schemes for one- and two-dimensional magnetohydrodynamics equations. II: High-order semidiscrete schemes.*, SIAM J. Sci. Comput. **28** (2006), no. 2, 533–560 (English).
3. Dinshaw S. Balsara, *Divergence-free adaptive mesh refinement for magnetohydrodynamics*, J. Comput. Phys. **174** (2001), 614–648.
4. Dinshaw S. Balsara and Daniel S. Spicer, *A staggered mesh algorithm using high order Godunov fluxes to ensure solenoidal magnetic fields in magnetohydrodynamic simulations*, J. Comput. Phys. **149** (1999), 270–292.
5. Nicolas Besse and Dietmar Kröner, *Convergence of locally divergence-free discontinuous-Galerkin methods for the induction equations of the 2D-MHD system*, M2AN Math. Model. Numer. Anal. **39** (2005), no. 6, 1177–1202.
6. Jeremiah U. Brackbill and D. C. Barnes, *The effect of nonzero $\nabla \cdot B$ on the numerical solution of the magnetohydrodynamic equations*, J. Comput. Phys. **35** (1980), 426–430.
7. Michael Breuß, *The correct use of the Lax-Friedrichs method*, M2AN Math. Model. Numer. Anal. **38** (2004), no. 3, 519–540.
8. *Clawpack (conservation laws package)*, <http://www.amath.washington.edu/claw>.
9. Constantine M. Dafermos, *Quasilinear hyperbolic systems with involutions*, Arch. Ration. Mech. Anal. **94** (1986), 373–389.
10. ———, *Hyperbolic conservation laws in continuum physics*, Springer, Berlin, Heidelberg, 2000.
11. Andreas Dedner, Friedemann Kemm, Dietmar Kröner, Claus-Dieter Munz, Thomas Schnitzer, and Matthias Wesenberg, *Hyperbolic divergence cleaning for the MHD equations*, J. Comput. Phys. **175** (2002), no. 2, 645–673.
12. Bernd Einfeldt, *On Godunov-type methods for gas dynamics.*, SIAM J. Numer. Anal. **25** (1988), no. 2, 294–318 (English).
13. C. R. Evans and John F. Hawley, *Simulation of general relativistic magnetohydrodynamic flows: A constrained transport method*, Astrophys. J. **332** (1988), 659.
14. F.G. Fuchs, S. Mishra, and N.H. Risebro, *Splitting based finite volume schemes for ideal MHD equations.*, J. Comput. Phys. **228** (2009), no. 3, 641–660 (English).
15. Franz G. Fuchs, Kenneth H. Karlsen, Siddharta Mishra, and Nils H. Risebro, *Stable upwind schemes for the magnetic induction equation.*, ESAIM, Math. Model. Numer. Anal. **43** (2009), no. 5, 825–852 (English).
16. Peter A. Gilman, *Magnetohydrodynamic “shallow water” equations for the solar tachocline*, Astrophys. J. Letters **544** (2000), no. 2, L79.
17. Amiram Harten, *High resolution schemes for hyperbolic conservation laws*, J. Comput. Phys. **49** (1983), no. 3, 357–393.
18. Amiram Harten and James M. Hyman, *Self adjusting grid methods for one-dimensional hyperbolic conservation laws*, J. Comput. Phys. **50** (1983), no. 2, 235–269.
19. Amiram Harten, Peter D. Lax, and Bram Van Leer, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev. **25** (1983), no. 1, 35–61.

20. Friedemann Kemm, *A carbuncle free Roe-type solver for the Euler equations.*, Benzoni-Gavage, Sylvie (ed.) et al., Hyperbolic problems. Theory, numerics and applications. Proceedings of the 11th international conference on hyperbolic problems, Ecole Normale Supérieure, Lyon, France, July 17–21, 2006. Berlin: Springer. 601-608 (2008), 2008.
21. ———, *Discrete involutions, resonance, and the divergence problem in mhd*, Hyperbolic Problems: Theory, Numerics and Applications (Eitan Tadmor, Jian-Guo Liu, and Athanasios E. Tzavaras, eds.), Proceedings of Symposia in Applied Mathematics, AMS, 2009, pp. 725–736.
22. ———, *A comparative study of TVD limiters – well known limiters and an introduction of new ones*, Internat. J. Numer. Methods Fluids (accepted), DOI: 10.1002/fld.2357.
23. Friedemann Kemm, Yong-Joong Lee, Claus-Dieter Munz, and Rudolf Schneider, *Divergence cleaning in finite volume computations for electromagnetic wave propagations*, Finite Volumes for Complex Applications – Problems and Perspectives (Raphaèle Herbin and Dietmar Kröner, eds.), 2002, pp. 575–582.
24. Randall J. LeVeque, *Finite volume methods for hyperbolic problems*, Cambridge University Press, Cambridge UK, 2002.
25. Barry Marder, *A method incorporating gauß’ law into electromagnetic pic codes*, J. Comput. Phys. **68** (1987), 48–55.
26. Siddharta Mishra and Eitan Tadmor, *Constraint preserving schemes using potential-based fluxes. i. multidimensional transport equations*, Preprint.
27. ———, *Constraint preserving schemes using potential-based fluxes. ii. genuinely multi-dimensional central schemes for systems of conservation laws*, Preprint.
28. ———, *Constraint preserving schemes using potential-based fluxes. iii. genuinely multi-dimensional schemes for the mhd equations*, Preprint.
29. Claus-Dieter Munz, Pascal Omnes, Rudolf Schneider, Eric Sonnendrücker, and Ursula Voss, *Divergence correction techniques for Maxwell solvers based on a hyperbolic model*, J. Comput. Phys. **161** (1999), 484–511.
30. Claus-Dieter Munz, Rudolf Schneider, and Ursula Voß, *A finite volume method for the Maxwell equations in the time domain*, SIAM J. Sci. Comput. **22** (2000), no. 2, 449–475.
31. Claus-Dieter Munz, Rudolf Schneider, Eric Sonnendrücker, and Ursula Voss, *Maxwell’s equations when the charge conservation is not satisfied*, C. R. Acad. Sci. Paris **328** (1999), no. S’erie I, 431–436.
32. Haim Nessyahu and Eitan Tadmor, *Non-oscillatory central differencing for hyperbolic conservation laws*, J. Comput. Phys. **87** (1990), 408–463.
33. Kenneth G. Powell, Phil L. Roe, R. Š. Myong, T. Gombosi, and D. de Zeeuw, *An upwind scheme for magnetohydrodynamics*, Workshop Methodes numeriques pour la M.H.D. (Luminy, France), 1995.
34. Kenneth G. Powell, Philipp L. Roe, Timur Linde, Tamas I. Gombosi, and Darren L. DeZeeuw, *A solution-adaptive upwind scheme for ideal magnetohydrodynamics*, J. Comput. Phys. **154** (1999), 284–309.
35. James A. Rossmannith, *An unstaggered high-resolution constrained transport method for magnetohydrodynamic flows*, SIAM J. Sci. Comput. **28** (2006), no. 5, 1766–1797.
36. Hans De Sterck, *Multi-dimensional upwind constrained transport on unstructured grids for “shallow water” magnetohydrodynamics*, AIAA 20012623, 2001.
37. Manuel Torrilhon, *Zur Numerik der idealen Magnetohydrodynamik*, Ph.D. thesis, ETH Zürich, 2003.
38. ———, *Locally divergence-preserving upwind finite volume schemes for magnetohydrodynamic equations*, SIAM J. Sci. Comput. **26** (2005), no. 4, 1166–1191.
39. Manuel Torrilhon and Michael Fey, *Constraint-preserving upwind methods for multidimensional advection equations*, SIAM J. Numer. Anal. **42** (2004), no. 4, 1694–1728.
40. Knut Waagan, *A positive muscl-hancock scheme for ideal magnetohydrodynamics*, J. Comput. Phys. **228** (2009), no. 23, 8609–8626.
41. Andrew L. Zachary, Andrea Malagoli, and Phillip Colella, *A higher-order Godunov method for multidimensional ideal magnetohydrodynamics*, SIAM J. Sci. and Stat. Comp. **15** (1994), no. 2, 263–284.

INSTITUTE FOR APPLIED MATHEMATICS AND SCIENTIFIC COMPUTING, BRANDENBURGISCHE TECHNISCHE UNIVERSITÄT COTTBUS, KONRAD-WACHSMANN-ALLEE 1, D-03046 COTTBUS, GERMANY

Current address: Department of Mathematics, Saarland University, Campus E2 4, D-66123 Saarbrücken, Germany

E-mail address: kemm@math.tu-cottbus.de

URL: <http://www.math.tu-cottbus.de/~kemm/>