



# Stability Issues of Entropy-Stable and/or Split-form High-order Schemes

## Analysis of Linear Stability

Gregor J. Gassner<sup>1</sup> · Magnus Svård<sup>3</sup> · Florian J. Hindenlang<sup>2</sup>

Received: 17 July 2020 / Revised: 15 October 2021 / Accepted: 3 November 2021 /  
Published online: 16 January 2022  
© The Author(s) 2022

### Abstract

The focus of the present research is on the analysis of *local energy stability* of high-order (including split-form) summation-by-parts methods, with e.g. two-point entropy-conserving fluxes, approximating non-linear conservation laws. Our main finding is that local energy stability, i.e., the numerical growth rate does not exceed the growth rate of the continuous problem, is not guaranteed even when the scheme is non-linearly stable and that this may have adverse implications for simulation results. We show that entropy-conserving two-point fluxes are inherently locally energy unstable, as they can be dissipative or anti-dissipative. Unfortunately, these fluxes are at the core of many commonly used high-order entropy-stable extensions, including split-form summation-by-parts discontinuous Galerkin spectral element methods (or spectral collocation methods). For the non-linear Burgers equation, we further demonstrate numerically that such schemes cause exponential growth of errors during the simulation. Furthermore, we encounter a similar abnormal behaviour for the compressible Euler equations, for a smooth exact solution of a density wave. Finally, for the same case, we demonstrate numerically that other commonly known split-forms, such as the Kennedy and Gruber splitting, are also locally energy unstable.

---

Gregor Gassner was supported by the European Research Council (ERC) under the European Union's Eight Framework Program Horizon 2020 with the research project Extreme, ERC Grant Agreement No. 714487.

---

✉ Gregor J. Gassner  
ggassner@uni-koeln.de  
Magnus Svård  
Magnus.Svard@uib.no  
Florian J. Hindenlang  
florian.hindenlang@ipp.mpg.de

<sup>1</sup> Department for Mathematics/Computer Science, Center for Data and Simulation Science, University of Cologne, Cologne, Germany

<sup>2</sup> Max-Planck Institute for Plasma Physics, NMPP division, Garching, Germany

<sup>3</sup> Department of Mathematics, University of Bergen, P.O. Box 7803, 5020 Bergen, Norway

**Keywords** Entropy-conserving two-point flux · Entropy-stable high-order scheme · Summation-by-parts · Split-form · Discontinuous Galerkin · Non-linear stability · Linear stability · Energy stability

## 1 Introduction

For consistent discretisations of well-posed linear partial differential equations (PDEs), it is well-known that stability (typically in  $L^2$ ) is necessary and sufficient for a numerical solution to converge to the analytical solution. This was proven by Lax and Richtmyer in [24]. However, many important PDEs are not linear and the Lax-Richtmyer convergence theorem is not valid. Examples include the Burgers equation and the Euler equations of compressible gas dynamics, which are of primary interest in the present work. It is equally well-known that the equations of fluid dynamics are not known to be well-posed and it is therefore not known what stability properties they have. Nevertheless, there is not a complete lack of knowledge. For smooth solutions of non-linear PDEs, stability of the (well-posed) linearised problem is sufficient to ensure convergence of the numerical solution with very fine grids as shown in [39] by Strang. That is, Strang's theorem reverts the problem back to Lax-Richtmyer's convergence theorem to prove linear well-posedness and stability for subsonic flows, where high-order accurate schemes have been proven to be very efficient, see e.g. [40]. (The necessity of stability for convergence will be discussed in Sect. 1.4.)

Next, we turn to solutions in the non-linear regime. For transonic or supersonic flows, shocks may appear. To capture shocks, an appropriate amount of artificial diffusion must be added to the scheme. This amounts to ensuring non-linear stability rather than linear. Nevertheless, once the scheme is stabilised, in the sense that the simulation does not blow up, it is again often assumed that the scheme is now convergent. However, for non-linear equations one bound (e.g. in  $L^2$ ) on the solution  $u$  is typically not enough to infer (even weak) convergence for the non-linear fluxes  $f(u)$ . For instance, in case of the Burgers equation (see Eq. (1) below), the flux is  $f(u) = u^2/2$  and if a sequence of approximate solutions,  $u^n \in L^2$ , converges weakly to  $u$ ,  $(u^n)^2$  does not necessarily converge to  $u^2$  (in  $L^1$ ), see e.g. [8]. In other words, a numerical solution may be perfectly stable and yet fail to converge to a solution. We also remark, that failure to converge may not only appear for solutions with shocks. Turbulence is a highly non-linear phenomena where linear stability may not be sufficient to infer convergence.

Since for the flow equations there is little mathematical guidance to what object we want our numerical schemes to converge, we turn to physics. A solution of the compressible Euler equations should not only satisfy the equations but also the second law of thermodynamics. That is, the total specific entropy should only increase in a closed system. This translates into the solution satisfying an extra partial differential inequality. Mathematically this extra condition implies some control (stability) of the solution. This has led to the development of so-called entropy-conservative/stable schemes, e.g., [1,6,11–13,17,19,20,28,32–34,36–38,41,42,47,48].

Herein, we study the local energy stability of entropy-conservative/stable high-order (including split-form) summation-by-parts (SBP) schemes. In the following subsections, we define the concepts of linear stability, entropy-stability and local energy stability. In particular, we differentiate between entropy-stability and entropy-dissipation.

### 1.1 Linear Stability

Consider a (system of) conservation law(s) on the periodic (in space) domain  $[-1, 1] \times (0, T]$ .

$$u_t + f(u)_x = 0, \quad u(x, 0) = u_0(x), \tag{1}$$

where  $u$  is the solution vector/scalar and  $f(u)$  is the vector/scalar flux function. The problem is closed with the initial datum  $u_0$ .

We begin by considering stability of linear problems. A linear PDE,  $u_t = Pu$ , on a periodic domain  $\Omega$  with initial data  $u_0 \in L^2([-1, 1])$ , is well-posed if there exists a unique solution  $\|u(\cdot, T)\|_2 = K_c \exp(\alpha_c T) \|u_0\|_2$  where  $\|u\|_2 = (\int_{\Omega} u^2 dx)^{1/2}$  is the  $L^2$  norm. The constants  $K_c$  and  $\alpha_c$  are independent of the initial data. (We tacitly assume that the estimate is sharp.) Let  $\mathbf{u}_t = P_h \mathbf{u}$  be a consistent approximation of the PDE, where  $\mathbf{u} = (u_0, u_1, \dots, u_N)^T$  is the discrete solution vector and

$$\|\mathbf{u}\|^2 = h\mathbf{u}^T \mathbf{u} = \sum_{i=0}^N hu_i^2 \tag{2}$$

an  $L^2$  equivalent norm. We say that the scheme is *stable*, if  $(\|\mathbf{u}(\cdot, T)\|_2)_t = K_n \exp(\alpha_n T) \|\mathbf{u}_0\|_2$ . We refer to [18] for more information on linear well-posedness and stability. Clearly, this notion of stability implies convergence in the sense of Lax-Richtmyer. When  $K_n$  and  $\alpha_n$  are any bounded constants, we term the scheme *Lax stable*. In view of Lax-Richtmyer’s Equivalence Theorem, convergence implies (Lax) stability. That is, Lax stability is *necessary* for convergence of a consistent scheme.

**Remark 1** We remind that the notions *stability* and *convergence* concern classes (e.g.  $L^2$ ) of initial data and solutions. A scheme that is neither stable nor convergent, need neither blow up nor fail to converge for every simulation.

To introduce some more restrictive stability notions, we consider the linear advection equation. That is, we consider (1) with  $f = au$  where  $a$  is the constant advection coefficient. By multiplying from the left by  $u$  and integrating in space and time, using the periodic boundary conditions, we obtain,  $\|u(\cdot, T)\| = \|u_0\|$ . The same estimate can be derived using Fourier analysis, and it is sharp. As is well-known, all eigenvalues of the symbol are purely imaginary. That is,  $\alpha_c = 0$ . If no low-order term ( $u$  without a derivative) is present in the PDE,  $\alpha_c = 0$  for any linearly well-posed PDE [18].

Turning to a semi-discretisation of (1) with the linear flux function, let  $u_i(t)$  be the approximation of  $u(x_i, t)$ , where  $x_i = ih, i = 0, \dots, N$  and  $h > 0$  is the grid spacing. Then we consider the semi-discrete finite volume scheme with a central flux,

$$(u_i(t))_t + \frac{f_{i+1/2} - f_{i-1/2}}{h} = 0, \tag{3}$$

$$f_{i+1/2}^{CN} = \frac{f(u_{i+1}) + f(u_i)}{2} = \frac{au_{i+1} + au_i}{2}. \tag{4}$$

By inserting (4) into (3), multiplying by  $hu_i$  and summing  $i$  from 0 to  $N$ , we obtain  $\|\mathbf{u}(\cdot, T)\| = \|\mathbf{u}_0\|$ , where  $\mathbf{u} = (u_0, u_1, \dots, u_N)^T$  with the discrete norm (2).

One typically refers to this notion as  $L^2$  stability, or energy well-posedness, or energy stability. As in the continuous case, the estimate can be arrived at by Fourier analysis and the eigenvalues of the discrete symbol are all imaginary. Either way, for the discretization with the central flux we have  $\alpha_n = \alpha_c = 0$ . The growth of the numerical scheme never exceeds that of the continuous equation.

Clearly, energy stability with  $\alpha_n = 0$  implies Lax stability (and convergence). However, Lax stability can also hold for a scheme with eigenvalues with positive real parts. (Positive eigenvalues, for short.) For instance, we could add an anti-dissipative term with a coefficient that vanishes as  $h \rightarrow 0$ . Such a scheme would have  $\alpha_n \sim h$  and would converge. One could even have  $\alpha_n = \text{constant}$ . Since we require the scheme to be consistent, those eigenvalues have to be associated with the highest frequency modes, which are small for  $L^2$  bounded data. However, Lax stable schemes are often impractical since they might require an excessive resolution to control the exponentially growing errors. For that reason, we want energy stability to guarantee that the growth of the numerical scheme can never exceed that of the continuous problem. As energy-stability implies Lax stability, we refer to linear stability if the scheme is at least Lax stable.

**Remark 2** A well-posed linear constant-coefficient PDE, with no zeroth-order term, has a spectrum with non-positive eigenvalues. For well-posed linear PDEs with variable coefficients and/or zeroth-order term the spectrum of the spatial differential operator is bounded, but possibly positive, and a Lax stable scheme may exceed the maximal growth of the continuous problem. See [18] for further information.

With these definitions of Lax stability, energy stability, and linear stability, the central difference scheme (3)-(4) is the marginally energy stable scheme.

### 1.2 Entropy-Stability

While entropy-stability analysis coincides with the linear stability analysis from the previous section in certain special cases (e.g. analysis of scalar linear problems with quadratic entropy), the main motivation for entropy-stability is the investigation of non-linear problems. We define an entropy (function) for a (non-linear) conservation law as a convex function  $U(u)$  that also satisfies  $U_u f_u = F_u$  where  $F(u)$  is the corresponding entropy flux.

A weak solution,  $u$ , of (1) is termed *entropy solution*, if it also satisfies,

$$U(u)_t + F(u)_x \leq 0, \tag{5}$$

which is an incarnation of the second law of thermodynamics.

The idea of entropy-stability is to design a numerical scheme for (1) that also satisfies (5). In the review [43], it is detailed how to achieve this for a finite volume scheme. We summarise the results here. First, we denote the vector  $U_u = w^T$  as the *entropy variables*. We consider the finite volume scheme (3) with non-linear flux functions  $f(u)$ . Multiplying by the entropy variables of cell  $i$ ,  $w_i^T$ , and assuming that,

$$(w_{i+1}^T - w_i^T) f_{i+1/2} - (\Psi_{i+1} - \Psi_i) \leq 0, \tag{6}$$

one arrives at,

$$(U_i)_t + \frac{F_{i+1/2} - F_{i-1/2}}{h} \leq 0, \quad \forall i, \tag{7}$$

where  $\Psi = w^T f - F$  is the *entropy potential* and  $F_{i+1/2} = \frac{1}{2}(w_{i+1}^T + w_i^T) f_{i+1/2} - \frac{1}{2}(\Psi_{i+1} + \Psi_i)$  is the numerical entropy flux. The inequality (7) is a consistent approximation of (5). What remains to do is to find fluxes  $f_{i+1/2}$  that satisfy (6).

As there is a somewhat confusing and loose definition of entropy-stability in the recent high-order literature, we introduce the following nomenclature used throughout this paper: A scheme is

- i. *entropy-stable*, if (7)/(6) is satisfied at all  $i$  and for *any* admissible entropy function  $U(u)$ ;
- ii. *entropy-conservative*, if (7)/(6) is satisfied as an equality for *one* entropy function at all  $i$ ;
- iii. *entropy-dissipative*, if it satisfies (7)/(6) but is neither entropy-stable nor entropy-conservative.

We will also refer to fluxes as having these properties with respect to (6) (locally).

**Remark 3** We refer to e.g. Tadmor [43] for a proof that both the Rusanov or Lax-Friedrichs schemes are entropy-stable discretisations. There are several ways to extend the finite volume entropy-conserving or dissipative scheme to high-order accuracy. A popular approach is to use high-order interpolation/reconstruction based on entropy-conservative fluxes within the blocks/elements, while the fluxes across the grid blocks/element interfaces can be either entropy-conservative or dissipative, see e.g. [2,3,10,11,14,26,27]. Note that using entropy-stable fluxes in the volume terms reduces the accuracy to first or second order.

### 1.3 Local Energy Stability

In fluid dynamics, e.g. in turbulence research [29], it is common to investigate the stability of certain flow states and their transition from a linear laminar behaviour to a non-linear turbulent behaviour. The analysis proceeds by linearising the non-linear system of conservation laws around a given baseflow state, e.g. [46]. The linearised equation reveals if the baseflow state is locally stable in space and time (the flow remains laminar), or allows a growth (local instability) and transitions to turbulence (non-linear behaviour).

This idea is closely connected to linear stability in the sense of Strang [39]. It is reasonable to demand that the numerical scheme has the same local behaviour as the equations themselves. For linear PDEs, we can immediately state that energy stable schemes have this property (they do not allow nonphysical growth, as the growth of the numerical scheme does not exceed the growth of the continuous problem), while Lax stable schemes might violate this principle.

Next, we apply the idea used in fluid dynamics research to extend the linear stability analysis to non-linear equations. Let us assume that we have a smooth 'baseflow'  $\tilde{u}(x, t)$  that satisfies the non-linear PDE. In hydrodynamics research, this state is typically a steady state, but here we allow a time dependence. We will also interchangeably denote the baseflow as the linearisation state throughout the manuscript. In general, local linear stability concerns the time evolution of small fluctuations  $|u'(x, t)| \ll |\tilde{u}(x, t)|$  added to the baseflow. Thus, we make the ansatz  $u(x, t) = \tilde{u}(x, t) + u'(x, t)$  for the solution of the non-linear problem and derive a linear equation for the fluctuations  $u'$ . Consider Burgers equation, i.e., (1) with  $f(u) = u^2/2$  on a periodic domain. We insert the ansatz

$$(\tilde{u}(x, t) + u'(x, t))_t + \frac{1}{2} ((\tilde{u}(x, t) + u'(x, t))(\tilde{u}(x, t) + u'(x, t)))_x = 0, \tag{8}$$

neglect all but the leading order differential terms (zeroth-order terms do not affect linear stability, see [18]), and use that  $\tilde{u}$  is an exact solution to arrive at

$$(u')_t + (\tilde{u}(x, t) u')_x = 0, \tag{9}$$

which is a variable coefficient advection equation in conservative form with the advection velocity  $\tilde{u}(x, t)$  for the fluctuations  $u'(x, t)$ . As mentioned above, we are interested in the stricter stability property, i.e., the local energy stability of the non-linear problem, at the

linearisation state  $\tilde{u}(x, t)$ , which is determined by the energy analysis of (9). Multiplying by  $u'$  and integrating over the domain, we get after some further manipulations

$$\frac{d}{dt} \|u'\|^2(t) = \int_{-1}^1 (\tilde{u}(x, t))_x (u'(x, t))^2 dx. \quad (10)$$

Since  $(\tilde{u})_x$  is a known and integrable function, we obtain

$$\|u'(\cdot, T)\|^2 \leq \exp \left[ T \frac{1}{2} \sup_{x,t} |(\tilde{u}(x, t))_x| \right] \|u'(\cdot, 0)\|^2. \quad (11)$$

This shows that in general, a bounded growth  $\alpha_c = \frac{1}{2} \sup_{x,t} |(\tilde{u}(x, t))_x| > 0$  of the fluctuations is possible. Hence, the variable coefficient advection problem is linearly well-posed. At the end of Sect. 2.2, we present a sharper estimate than (11) for a specific choice of baseflows  $\tilde{u}$ , i.e., for a specific class of variable coefficient problems, that enables us to identify and rule out schemes that are only Lax stable.

#### 1.4 General Discussion on Stability and Convergence

We end the Introduction with a general discussion on stability and convergence. As already mentioned, if a problem is ( $L^2$ -)well-posed, a convergent scheme should converge for any  $L^2$  bounded initial data. Non-convergence does not imply that solutions *always* diverge. The same is true for *stable* vs. *unstable* schemes.

Turning to a non-linear problem, Strang's theorem states that linear stability (either the stricter energy stability or the weaker Lax stability) of the first variation is sufficient (and necessary since a convergent sequence is bounded) for convergence to a smooth solution. The smoothness assumption guarantees that the first variation dominates the higher-order variations for sufficiently small perturbations. However, the first variation is a variable coefficient problem with zeroth-order terms, which may induce a growth. Linear stability guarantees that any growth is bounded, as seen in Sect. 1.1. Indeed, this growth may push the solution into a regime where the non-linear terms are no longer negligible. This is the mechanism that produces turbulence or shocks even when energy stable, such as central difference, schemes are used. (In the same way, artificial growth that exceeds the growth of the continuous problem caused by an energy unstable operator may erroneously push the solution into the non-linear regime.) Hence, the spectrum of an energy stable difference operator approximating a variable coefficient problem may be positive (but the real part is bounded from above). Since variable coefficients and zeroth-order terms are "benign" with respect to linear stability, they are often ignored and the problem is reduced to the constant coefficient case (known as "freezing" the coefficients.) The constant coefficient case highlights the effect of the difference operators and conveniently makes the continuous spectrum non-positive. An energy stable scheme mimics this property, while a Lax stable scheme may allow growth in the constant coefficient problem. In the cases we study numerically, we do not freeze the coefficients and to be able to compare schemes, we therefore choose continuous problems without growth, i.e., a special case where  $\alpha_c = 0$ . (The theory underpinning these statements is found in [18].)

The remainder of the paper is organised as follows: In Sect. 2, we present the local stability analysis of entropy-conservative/dissipative schemes for Burgers equation, including skew-symmetric high-order discontinuous Galerkin spectral element approximations. The analysis

reveals local energy stability issues with such schemes and that they are Lax stable at most. In Sect. 3, we show that the local energy stability issues carry over to the case of the compressible Euler equations in multi-dimensions and provide further numerical evidence. We draw our conclusions in Sect. 4.

## 2 Local Energy Stability Analysis of the Non-linear Burgers Equation

The split-form of the non-linear Burgers equation is

$$u_t + \alpha (u^2/2)_x + (1 - \alpha) u u_x = 0, \tag{12}$$

where  $\alpha \in [0, 1]$  is the split parameter. The choice  $\alpha = 1$  gives the divergence form of the PDE, whereas  $\alpha = 0$  gives the advective quasi-linear form. It is well-known that the 2/3 - trick can be used to derive an entropy estimate for the non-linear equation. The Burgers equation with  $\alpha = 2/3$  takes the form

$$u_t + \frac{1}{3} u u_x + \frac{1}{3} (u^2)_x = 0. \tag{13}$$

Multiplying by  $u$  and integrating in space gives

$$\frac{1}{2} (\|u\|^2)_t + \left[ \frac{1}{3} (u^3) \right]_{-1}^1 = 0, \tag{14}$$

and with periodicity

$$\frac{d}{dt} (\|u\|^2) = 0. \tag{15}$$

This trick is used to design entropy-conservative fluxes and entropy-conserving split-form schemes.

### 2.1 The Finite Volume Perspective

By approximating (12), with central finite differences, we obtain

$$(u_i)_t + \alpha \frac{\frac{1}{2}u_{i+1}^2 - \frac{1}{2}u_{i-1}^2}{2h} + (1 - \alpha) u_i \frac{u_{i+1} - u_{i-1}}{2h} = 0. \tag{16}$$

It is straight forward to check that the approach is conservative, as it can be recast into the finite volume form (3) with the split-flux function

$$f_{i+1/2}^\alpha = \frac{1}{2} (f(u_i) + f(u_{i+1})) - \frac{1}{2} \frac{(1 - \alpha) (u_{i+1} - u_i)}{2} (u_{i+1} - u_i). \tag{17}$$

The choice  $\alpha = 2/3$  corresponds to the skew-symmetric form (13) and gives the entropy-conserving flux function

$$f_{i+1/2}^{EC} = \frac{1}{2} (f(u_i) + f(u_{i+1})) - \frac{1}{2} \left( \frac{u_{i+1} - u_i}{6} \right) (u_{i+1} - u_i) = \frac{1}{6} (u_i^2 + u_i u_{i+1} + u_{i+1}^2). \tag{18}$$

To gain some further insights into the properties of the fluxes (17) and (18), we interpret them as a standard finite volume flux of the form,

$$f_{i+1/2} = \frac{1}{2} (f(u_i) + f(u_{i+1})) - \frac{1}{2} R_{i+1/2}(u_{i+1} - u_i), \tag{19}$$

where  $R_{i+1/2}$  is a dissipation coefficient/matrix. We hence get for the fluxes (17) and (18) the dissipation coefficients

$$R_{i+1/2}^\alpha = \frac{(1 - \alpha) (u_{i+1} - u_i)}{2}, \quad R_{i+1/2}^{EC} = \left( \frac{u_{i+1} - u_i}{6} \right). \tag{20}$$

Note, that the coefficients (20) are not necessarily definite. Only for the choice  $\alpha = 1$ , the dissipation coefficient is guaranteed to be zero for all values  $u_{i+1}$ ,  $u_i$ , and the fluxes reduce to the central flux. For general  $\alpha$ , in particular for the skew-symmetric entropy-conserving choice  $\alpha = 2/3$ , the flux can be positive and thus dissipative, but also negative and thus anti-dissipative. We note that for the entropy-conserving flux, this was already observed in [43].

By linearising the finite volume scheme (3) with the split-fluxes (17) around a smooth baseflow  $\tilde{u}$  that is assumed to satisfy the scheme, it is straightforward to deduce that we get again a finite volume form (3) with the linearised fluxes

$$\tilde{f}_{i+1/2}^\alpha = \frac{(\tilde{u}_{i+1} u'_{i+1}) + (\tilde{u}_i u'_i)}{2} - \frac{1}{2} \tilde{R}_{i+1/2}^\alpha (u'_{i+1} - u'_i), \tag{21}$$

where the dissipation coefficient of the linearised flux is

$$\tilde{R}_{i+1/2}^\alpha = (1 - \alpha) (\tilde{u}_{i+1} - \tilde{u}_i). \tag{22}$$

We note that the linearised flux  $\tilde{f}_{i+1/2}^\alpha$  is consistent with the linearised problem (8) for all split parameters  $\alpha$ . Furthermore, the dissipation coefficient of the linearised fluxes  $\tilde{R}_{i+1/2}^\alpha$  are for arbitrary  $\alpha \neq 1$  positive or negative, depending on the slope of the baseflow  $\tilde{u}$ . Clearly, the scheme is stable for all  $\alpha$  in the sense of Lax for smooth  $\tilde{u}$ , but the anti-dissipative parts will cause positive eigenvalues and we thus do not have energy stability (see Sect. 1.1) and consequently the scheme is not locally energy stable.

We emphasise that the local energy analysis simply demonstrates that the anti-dissipative character observed in (20), is unchanged in the linear regime. Only for the choice  $\alpha = 1$  the dissipation coefficients are  $R_{i+1/2}^\alpha = 0$  and  $\tilde{R}_{i+1/2}^\alpha = 0$  for all  $i$ , which is the marginal case for energy stability. This is well-known since the scheme turns into the central difference scheme of the divergence form, which has purely imaginary eigenvalues.

**Remark 4** We note that for the entropy-conserving case,  $\alpha = 2/3$ , Tadmor [43] proposed to remove the anti-dissipative character by replacing the entropy-conserving dissipation coefficient  $R_{i+1/2}^{EC}$  in (20) with

$$R_{i+1/2}^{TA} = \max \left( \frac{1}{6} (u_{i+1} - u_i), 0 \right). \tag{23}$$

Then the corresponding numerical flux becomes entropy-dissipative with respect to the quadratic entropy  $U(u) = u^2/2$ .

An alternative way to construct an entropy-dissipative flux is to add an explicit viscosity term to an entropy-conserving flux. For instance, a Rusanov-type viscosity term of the form

$$\begin{aligned}
 f_{i+1/2}^{\text{ED}} &= \frac{1}{6}(u_i^2 + u_i u_{i+1} + u_{i+1}^2) - \frac{1}{2} \max(|u_i|, |u_{i+1}|) (u_{i+1} - u_i) \\
 &= \frac{1}{2} (f(u_i) + f(u_{i+1})) - \frac{1}{2} \left( \frac{u_{i+1} - u_i}{6} + \max(|u_i|, |u_{i+1}|) \right) (u_{i+1} - u_i),
 \end{aligned}
 \tag{24}$$

which corresponds to  $R_{i+1/2}^{\text{ED}} = \left( \frac{u_{i+1} - u_i}{6} + \max(|u_i|, |u_{i+1}|) \right) \geq 0$ . In contrast, the standard Rusanov flux function, see e.g. [44], is an entropy-stable flux function

$$f_{i+1/2}^{\text{ES}} = \frac{1}{2} (f(u_i) + f(u_{i+1})) - \frac{1}{2} \max(|u_{i+1}|, |u_i|) (u_{i+1} - u_i),
 \tag{25}$$

which corresponds to  $R_{i+1/2}^{\text{ES}} = \max(|u_{i+1}|, |u_i|)$ .

### 2.2 The Finite Difference Perspective

To give an alternative perspective on the issue, we consider the split-form (16) directly in its finite difference form. One reason of the popularity of the split-form schemes (16) is that for the choice  $\alpha = 2/3$  (skew-symmetric form) the continuous estimate (15) carries over

$$\frac{d}{dt} (\|\mathbf{u}\|^2) = 0.
 \tag{26}$$

Hence, the skew-symmetric scheme is neutrally stable in  $L^2$  and (once again) we conclude that it is *entropy-conservative* with respect to the quadratic entropy  $U(u) = u^2/2$ . This appears to be at odds with the conclusion above that the scheme is not locally energy stable. However, local energy stability refers to the local behaviour in space and time deduced from the linearised problem. The estimate (26) shows that there is an upper limit to the local (un-) physical growth. Below, however, we will demonstrate numerically, that this global bound allows for a significant growth of local fluctuations.

To analyse the finite difference approximation (16) we first rewrite it in the following equivalent way

$$(u_i)_t + \frac{\frac{1}{2}u_{i+1}^2 - \frac{1}{2}u_{i-1}^2}{2h} - \frac{(1-\alpha)}{2} h^2 \frac{(u_{i+1} - u_{i-1})}{2h} \frac{(u_{i+1} - 2u_i + u_{i-1})}{h^2} = 0.
 \tag{27}$$

We can reinterpret this form as a consistent discretisation of the following continuous problem

$$u_t + \left( \frac{u^2}{2} \right)_x - v u_{xx} = 0,
 \tag{28}$$

where the last term corresponds to dissipation with the coefficient

$$v = \frac{(1-\alpha)}{2} h^2 u_x,
 \tag{29}$$

that analogously to the finite volume perspective gets anti-dissipative for negative slopes  $u_x < 0$ , if  $\alpha \neq 1$  and  $h > 0$ .

As before, we linearise the scheme around a baseflow  $\tilde{u}_i(t)$  (which solves (16)) and insert the ansatz  $u_i = \tilde{u}_i + u'_i$  into the scheme (16). Neglecting high-order terms of the small

amplitude fluctuations, such as  $(u'_i)^2$ , we get

$$(u'_i)_t + \alpha \frac{\tilde{u}_{i+1}u'_{i+1} - \tilde{u}_{i-1}u'_{i-1}}{2h} + (1 - \alpha) \left[ \tilde{u}_i \frac{u'_{i+1} - u'_{i-1}}{2h} + u'_i \frac{\tilde{u}_{i+1} - \tilde{u}_{i-1}}{2h} \right] = 0, \tag{30}$$

which is a consistent approximation of

$$(u')_t + \alpha (\tilde{u} u')_x + (1 - \alpha) (\tilde{u} (u')_x + u' \tilde{u}_x) = 0, \tag{31}$$

which in turn is a split-form of the continuous linearised variable coefficient advection problem (9). Again, it is possible to equivalently rewrite this discretisation into

$$(u'_i)_t + \frac{\tilde{u}_{i+1}u'_{i+1} - \tilde{u}_{i-1}u'_{i-1}}{2h} - \frac{(1 - \alpha)}{2} h^2 \left[ \frac{(\tilde{u}_{i+1} - \tilde{u}_{i-1}) (u'_{i+1} - 2u'_i + u'_{i-1})}{2h} \frac{1}{h^2} + \frac{(u'_{i+1} - u'_{i-1}) (\tilde{u}_{i+1} - 2\tilde{u}_i + \tilde{u}_{i-1})}{2h} \frac{1}{h^2} \right] = 0, \tag{32}$$

which we can reinterpret as a consistent discretisation of the following continuous problem

$$u_t + (\tilde{u} u')_x - \frac{(1 - \alpha)}{2} h^2 (\tilde{u}_x u'_{xx} + u'_x \tilde{u}_{xx}) = 0 \tag{33}$$

$$u_t + (\tilde{u} u')_x - (\tilde{v} u'_x)_x = 0,$$

where the last term corresponds to dissipation with the coefficient

$$\tilde{v} = \frac{(1 - \alpha)}{2} h^2 \tilde{u}_x. \tag{34}$$

Analogously to the finite volume perspective, the coefficient gets anti-dissipative for negative slopes of the baseflow  $\tilde{u}_x < 0$ , if  $\alpha \neq 1$  and  $h > 0$ . Note that this may lead to a significant growth for a finite  $h$ . By taking  $h$  sufficiently small, the growth can be reduced but it may require extra resolution.

**Remark 5** The analysis was performed for general  $\alpha$  and shows that any split-form is at most Lax stable, except for the central approximation with  $\alpha = 1$ . In particular, the skew-symmetric split-form, that is entropy-conserving, is at most Lax stable.

The effect of split-form approximations for variable coefficient problems has also been studied in Manzanero et al. [30]. They showed that with the additional assumption  $\tilde{u}(x, t) > 0, \forall x, t$ , a sharper estimate than (11) can be derived

$$\min_{x \in [-1, 1]} \{ \tilde{u}(x) \} \|u'(\cdot, T)\|^2 \leq \max_{x \in [-1, 1]} \{ \tilde{u}(x) \} \|u'_0(\cdot)\|^2. \tag{35}$$

The estimate is interesting, as it shows that for the special choice of positive linearisation states there is no exponential growth of the solution, i.e.  $\alpha_c = 0$  in this particular case. In addition, Manzanero et al. showed that for positive linearisation states the central approximation based on the divergence form ( $\alpha = 1$ ) of the PDE with central fluxes is *the only scheme* that has the correct behaviour, i.e. there is no artificial exponential growth or damping. *The central scheme (3) with the numerical flux  $f_{i+1/2}^{CN} = \frac{1}{2}(f(u_i) + f(u_{i+1}))$  is thus the least diffusive energy stable scheme for such a problem.* This is exactly the same conclusion that we can draw based on the above analysis, where we interpreted the scheme as a central scheme plus (anti-)dissipation. The estimate (35) is very convenient for the later numerical experiments as it will allow us to identify and rule out Lax stable approximations and we will refer to it as the *MRFVK-estimate* (the initials of the authors).

**Remark 6** We note that it is possible to design either energy stable schemes for the linear problem, or entropy-dissipative/conservative schemes for the non-linear problem. However, it seems that the core of the issue is that discretisation of the non-linear PDE and linearisation of the resulting scheme is not the same as linearisation of the PDE and then discretising the linearised problem.

In view of the above observations, we will use the lack of local energy stability as an indicator that a scheme is anti-dissipative. There are two options to study the linearised equations: (1) Derive an energy estimate. Only anti-diffusive terms can cause a growth of the norm in a periodic linear scheme in the absence of low-order terms. (2) Compute the eigenvalues of the spatial operators. We consider problems that satisfy the MRFVK-estimate, i.e., with no exponential growth and  $\alpha_c = 0$ , such that we know a priori that the spectrum must have non-positive eigenvalues. Hence, we can determine if the scheme has anti-dissipative terms by examining if the eigenvalues have positive real parts.

### 2.3 Analysis of High-order Diagonal-norm Split-form Summation-by-parts Operators

In this subsection we focus on high-order diagonal-norm split-form summation-by-parts approximations with different entropic properties for the Burgers equation. In particular, we analyse the skew-symmetric discontinuous Galerkin spectral element method [14] that is strongly related to split-form summation-by-parts finite difference approximations with simultaneous approximation terms, e.g. [10]. We note that the entropy-conserving flux function is a key building block for the construction of high-order entropy-conserving/dissipative schemes, see e.g., [2,3,10,11,14,27].

Details on the high-order discretisation can be found e.g. in [14,17]. We give a short summary of the necessary building blocks. We split the computational domain into elements of equal size  $h$ . We map the elements to a unit reference element  $[-1, 1]$ . In each reference element, the solution is approximated by a polynomial of degree  $N$ . We use  $N + 1$  Legendre-Gauss-Lobatto nodes to represent the polynomials with Lagrangian basis functions. Thus, the solution is represented as a vector of nodal values  $\underline{u}$ . For convenience we also introduce the diagonal injection of the nodal values to simplify notation, i.e.  $\underline{U} = \text{diag}(\underline{u})$ . The corresponding main operators are the mass matrix  $\underline{M} = \text{diag}([\omega_0, \dots, \omega_N])$  (or integration matrix), the differentiation matrix  $\underline{D}$  and the boundary matrix  $\underline{B} = \text{diag}([-1, 0, \dots, 0, 1])$ . The  $\omega_i$  are the corresponding  $N + 1$  Legendre-Gauss-Lobatto integration weights in the reference element. The operators satisfy the summation-by-parts property

$$\underline{M} \underline{D} + (\underline{M} \underline{D})^T = \underline{B}, \tag{36}$$

where the mass matrix is interpreted as the diagonal-norm operator. We consider the split-form discontinuous Galerkin approximation

$$\frac{h}{2} \underline{u}_t + \alpha \frac{1}{2} \underline{D} \underline{U} \underline{u} + (1 - \alpha) \underline{U} \underline{D} \underline{u} = -\underline{M}^{-1} \underline{B} \left[ \underline{f}^* - \frac{1}{2} \underline{U} \underline{u} \right], \tag{37}$$

which follows directly from the continuous split-form (12) of the Burgers equation. The numerical flux vector is

$$\underline{f}^* = [f_{i-1/2}, 0, \dots, 0, f_{i+1/2}]^T, \tag{38}$$

where  $f_{i+1/2}$  is a consistent and unique numerical flux function, e.g. as used in the finite volume approach, see Sect. 2.1.

We get an entropy-conserving semi-discretisation with the choice  $\alpha = 2/3$  (skew-symmetric form) and  $\underline{f}^*$  taken to be the entropy-conserving numerical flux (18) for the quadratic entropy  $U(u) = u^2/2$ . We obtain entropy-dissipation when switching the numerical flux from entropy-conserving to entropy-dissipative, e.g. (24). Thus, these versions of the high-order discretisation have a global bound on the discrete  $L^2$ -norm, see e.g. [14].

Having defined the scheme, we proceed with the local energy stability analysis. As before, we consider a perturbation of a smooth baseflow  $\underline{u} = \tilde{\underline{u}} + \underline{u}'$  and assume that  $\tilde{\underline{u}}$  solves (37). The analysis of the numerical flux function  $f_{i+1/2}$  appearing in the surface integral, is similar to the finite volume discussion in Sect. 2.1. Similarly, if we linearise the volume terms by inserting the ansatz with the fluctuations, neglecting high-order fluctuation terms, we obtain

$$\begin{aligned} & \left[ \alpha \frac{1}{2} \underline{D} \underline{U} \underline{u} + (1 - \alpha) \underline{U} \underline{D} \underline{u} \right] - \left[ \alpha \frac{1}{2} \underline{D} \tilde{\underline{U}} \tilde{\underline{u}} + (1 - \alpha) \tilde{\underline{U}} \underline{D} \tilde{\underline{u}} \right] \\ & \Rightarrow \alpha \left[ \underline{D} \tilde{\underline{U}} \underline{u}' \right] + (1 - \alpha) \left[ \tilde{\underline{U}} \underline{D} \underline{u}' + \underline{U}' \underline{D} \tilde{\underline{u}} \right]. \end{aligned} \tag{39}$$

The analysis of these volume terms is analogous to the finite difference discussion in Sect. 2.2. As mentioned already before, for  $\tilde{\underline{u}} > 0$  (for all elements), the MRFVK-estimate (35) applies. The corresponding discrete estimate is only obtained when  $\alpha = 1$ , which is the central approximation with divergence form volume terms. All other choices, including the  $\alpha = 2/3$  entropy-conserving/dissipative choice, allow local artificial exponential growth of fluctuations, i.e., *they are not locally energy stable*.

We note that while the central approximation is locally energy stable, its deficiency is in the non-linear regime as the central approximation does not have a non-linear global bound. Hence, the central approximation is not preferable either.

### 2.4 Numerical Investigations for the Non-linear Burgers Equation

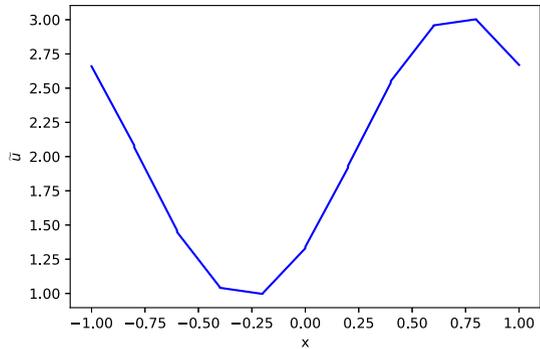
To numerically investigate our theoretical findings of the local energy stability issues, we consider the non-linear Burgers equation with periodic boundary conditions. To apply the MRFVK-estimate, we choose a smooth initial data that is positive in the domain

$$\tilde{\underline{u}}(x) = \sin(\pi x - 0.7) + 2, \quad x \in [-1, 1], \tag{40}$$

which will also serve as the linearisation state. As the corresponding linearised variable coefficient problem satisfies the MRFVK-estimate, there is no (local in time) exponential growth of the solution for this test problem and hence the maximum real part of the eigenvalues is zero, i.e.  $\alpha_c = 0$ . As discussed in the Sects. 2.1 and 2.2, positive eigenvalues of the spectrum of the linearised spatial operator for this particular test case indicate exponential growth exceeding the growth of the continuous problem (artificial anti-dissipation) and that the scheme is Lax stable at most.

Figure 1 shows an illustration of the baseflow for the investigations used below. To mimic mild under-resolution of the baseflow, we project the linearisation state (40) onto piece-wise linear polynomial basis functions ( $\tilde{N} = 1$ ) for all Burgers simulation results presented below. As can be seen in Fig. 1, the baseflow is discontinuous across element interfaces, although the jumps are relatively small as the baseflow is only mildly under-resolved.

**Fig. 1** Plot of the baseflow (40) for a discretisation with 10 elements and polynomial degree  $\tilde{N} = 1$  in every element. Note that there are small jumps at the element interfaces



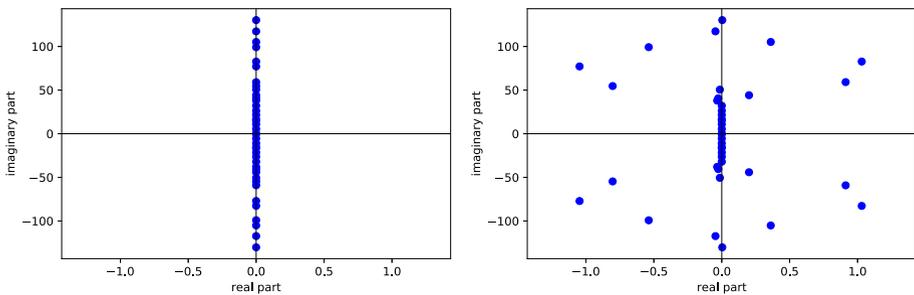
### 2.4.1 Comparison of the Spectra of the Semi-discrete Operator

To compute the spectrum of a spatial operator, we use the typical finite difference approach to approximate the linearisation and determine the  $j$ -th column of the Jacobi matrix of the semi-discrete residual as

$$Ae_j \approx \frac{\text{rhs}(\tilde{u} + e_j \epsilon) - \text{rhs}(\tilde{u} - e_j \epsilon)}{2\epsilon} \tag{41}$$

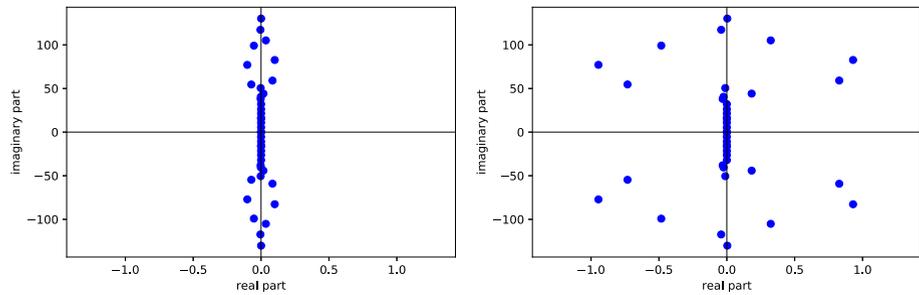
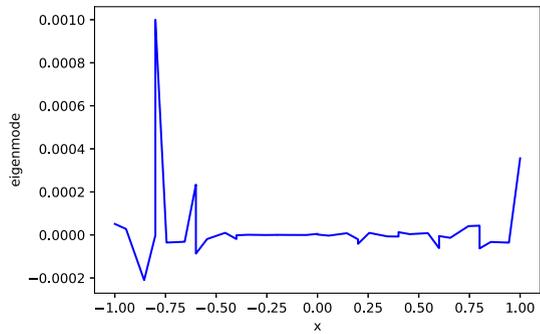
where  $\epsilon = 10^{-8}$  and  $e_j$  is the unit vector in the  $j$ -th direction. With the finite difference approach, the accuracy of the eigenvalues are limited to about single machine precision and hence, we interpret eigenvalues within this accuracy constraint as zero, even when formally the number is positive. The computed eigenvalues of the approximative Jacobi matrix  $A$  give an approximate spectrum of the semi-discrete high-order operator. We are particularly interested in assessing the maximum real part across all eigenvalues as positive eigenvalues imply exponential growth of their eigenmodes.

For the approximation (37), we plot the results of the fully central approximation ( $\alpha = 1$ ) and the entropy-conserving approximation ( $\alpha = 2/3$ ) in Fig. 2. The plots reveal a significant positive real part of 1.0307 for the entropy-conserving approximation, whereas the spectrum of the central approximation is neutrally stable and almost purely imaginary in accordance with (35).



**Fig. 2** Spectra of spatial operators. Left: Central approximation with the divergence form ( $\alpha = 1$ ) and the central numerical flux. Right: Entropy-conserving approximation with skew-symmetric form ( $\alpha = 2/3$ ) and the entropy-conserving flux (18). [Discretisation with polynomial degree  $N = 3$  and 10 elements with the baseflow shown in Fig. 1. The maximum real part of the approximate spectrum is about  $8.8 \times 10^{-8}$  for the central scheme and about 1.0307 for the entropy-conserving scheme]

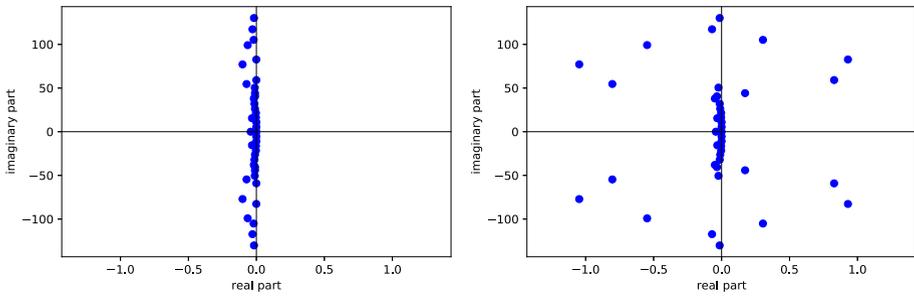
**Fig. 3** Plot of the eigenmode corresponding to the eigenvalue with the largest real part (about 1.0307) for the entropy-conserving discretisation with skew-symmetric form ( $\alpha = 2/3$ ) and the entropy-conserving flux (18). [The discretisation uses 10 elements and a polynomial degree  $N = 3$ , with the baseflow (40). The eigenmode is scaled such that its largest peak is 0.001]



**Fig. 4** Spectra of spatial operators. Left: Volume term in divergence form ( $\alpha = 1$ ), surface term uses entropy-conserving flux (18). Right: Volume term in skew-symmetric form ( $\alpha = 2/3$ ), surface term uses central flux. [Discretisation with polynomial degree  $N = 3$  and 10 elements with the baseflow shown in Fig. 1. The maximum real part is about 0.1006 (left) and 0.9300 (right)]

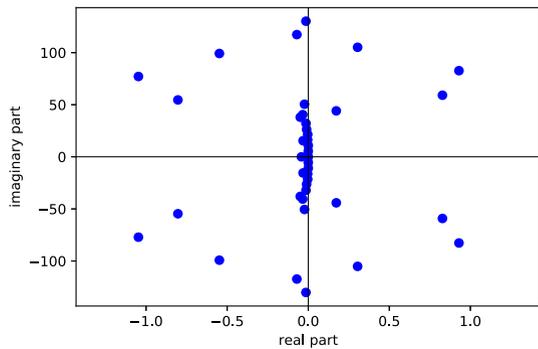
With a maximum real part of about 1.0307 in the defective spectrum, we should expect a strong exponential growth of the corresponding eigenmode. We plot the corresponding eigenmode in Fig. 3. We note that the eigenmode is active in the part of the domain, where the slope of the baseflow is negative. Although unclear if there is a direct connection, it is interesting to note that this coincides with the discussion on the entropy-conserving split-form in the previous Sect. 2.2, where the dissipation coefficient gets negative, when the slope of the linearisation state is negative.

Next, we investigate if either the volume terms or the surface terms are the cause for this faulty and undesired behaviour of the discretisation. Figure 4 shows the result for the divergence form of the volume terms ( $\alpha = 1$ ) combined with the entropy-conserving flux (18) in the left plot. In the right plot, the discretisation ingredients are flipped with skew-symmetric ( $\alpha = 2/3$ ) volume terms and the central numerical flux for the surface terms. It is clearly seen that the spectra still have eigenvalues with large positive real parts, here the maximum is 0.1006 in the left plot and 0.93 in the right plot. The exponential growth rate with the entropy-conserving surface flux is smaller than for the fully entropy-conserving scheme. The reason is, that the baseflow is relatively well resolved, with only small jumps at element interfaces. As the anti-dissipation scales directly with this jump, its effect is decreased. The plot on the right also suggests that the volume terms in entropy-conserving form are indeed a major contributor to the lack of local energy stability.



**Fig. 5** Spectra of spatial operators. Left: Volume term in divergence form ( $\alpha = 1$ ), surface term uses the modified entropy-stable flux that excludes anti-dissipation (23). Right: The volume terms are in skew-symmetric form ( $\alpha = 2/3$ ), the surface term uses the modified entropy-dissipative flux that excludes anti-dissipation (23). [Discretisation with polynomial degree  $N = 3$  and 10 elements with the baseflow shown in Fig. 1. The maximum real part is about  $-9.01 \times 10^{-8}$  (left) and 0.9298 (right)]

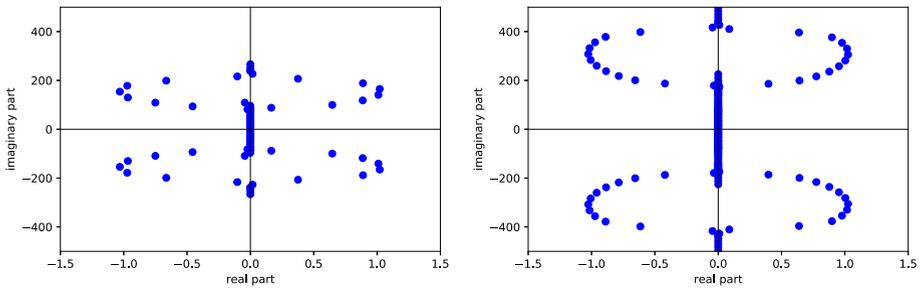
**Fig. 6** Spectrum of spatial operator. The volume terms are entropy-dissipative according to the modification of Carpenter et al. (see [2], Sect. 5, Eq. (5.6)). The surface flux is the modified version of Tadmor (23), that excludes anti-dissipation. [Discretisation with polynomial degree  $N = 3$  and 10 elements with the baseflow shown in Fig. 1. The maximum real part is about 0.9297]



We further investigate the impact of the modified entropy-conserving flux by Tadmor (23) that excludes the anti-dissipation. To demonstrate that it has a stabilising effect on the scheme, we redo the plots of Fig. 4. The results are illustrated in Fig. 5 and suggest that indeed Tadmor’s modification seems to get rid of the faulty influence of the anti-dissipation (left plot), when combined with divergence form volume terms. This corroborates our theoretical findings that the positive part of the spectrum is associated with the anti-diffusive character of the scheme. However, it is also clear, that this particular modification of the surface term is not enough to fix the locally energy unstable behaviour of the skew-symmetric volume terms, as can be seen in the right plot, which still has a significant positive real part, due to the anti-diffusion in the volume terms.

**Remark 7** It is interesting to note that Carpenter et al. considered a modification of the volume terms to introduce entropy-dissipation, where they use entropy generation of the central approximation compared to the entropy-conserving approximation to adjust the volume terms (see [2], Sect. 5, Eq. (5.6)). Unfortunately, our numerical investigations show only little influence on the spectra for this modification, as can be seen in Fig. 6.

In particular, we still observe significant positive real parts of about 0.9297, similar to the case with full skew-symmetric volume terms. This highlights that a sufficient amount of dissipation is necessary to pull the spectrum to the left-hand side.

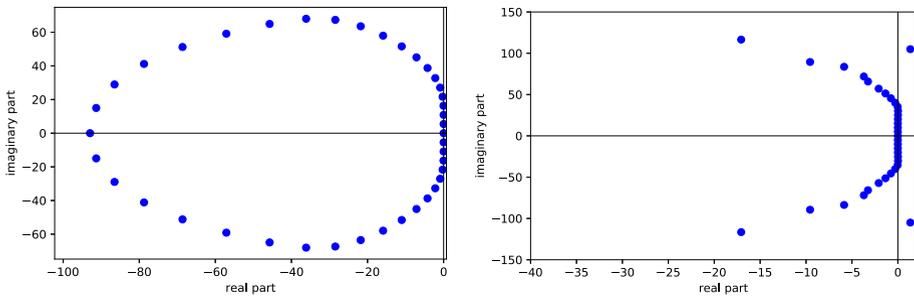


**Fig. 7** Spectra of spatial operator, for entropy-conserving approximation with skew-symmetric form ( $\alpha = 2/3$ ) and the entropy-conserving flux (18). Left: Number of elements is 20. Right: Number of elements is 40. [Discretisation with polynomial degree  $N = 3$ . Linearisation with the baseflow shown in Fig. 1. Maximum real part is 1.021 (left) and 1.025 (right)]

Next, we investigate the impact of resolution on the behaviour of the spectrum. As discussed in the previous Sect. 2.2, we are able to interpret the skew-symmetric forms as central discretisations with dissipation. The dissipation coefficient can be positive or negative and scale with the mesh size  $h$ , which suggests that refining the mesh (decreasing  $h$ ) should improve the faulty behaviour. We plot in Fig. 7 the resulting spectra of the entropy-conserving skew-symmetric form when increasing the number of elements from 10 (compare right part of Fig. 2) to 20 and 40 respectively. One can clearly see that the maximum positive real part does not shrink and stays at the value of about 1. However, we can also clearly observe that the faulty eigenmodes get pushed to higher frequencies, as the corresponding imaginary parts increase with decreasing mesh size. We note that we observe a similar behaviour for all discretisations discussed in this work: the faulty eigenmodes always get pushed to higher frequencies when refining the mesh, whereas the maximum real part stays about the same. These numerical investigations suggest that the high-order schemes might be Lax stable. Depending on the initial condition, we expect that for very well resolved smooth problems, the content of high frequency eigenmodes decays quickly. Thus, the contribution of the faulty high frequency eigenmodes should also be reduced when refining the mesh, compare the discussion at the end of Sect. 1.1, and convergence could be achieved. We refer to Sect. 2.4.2 for further numerical results that support this expectation, see the results in Figs. 13 and 14.

Lastly, we investigate the effect of a strongly dissipative numerical surface flux function. As in most practical simulations, the skew-symmetric volume terms are combined with a dissipative numerical surface flux, such as e.g. the Rusanov-type entropy-dissipative flux (24). From a theoretical point of view, it is not obvious how a surface dissipation term can apply control over volume type anti-dissipation. However, as can be seen in the left part of Fig. 8, the spectrum gets shifted towards the negative real axis, with the maximum real part of now  $-1.06 \times 10^{-7}$ .

Unfortunately, it is possible to design a discretisation setup where the surface dissipation is not enough to stabilise the skew-symmetric volume integrals. To provoke such a behaviour, we first increase the polynomial degree from  $N = 3$  to  $N = 15$  and decrease the number of elements from 10 to 3. This gives about the same number of degrees of freedom (48 in comparison to 40), however the number of element interfaces is smaller, which should obviously decrease the stabilising effect of the surface terms. We note that it is especially the high-order polynomial approximations that are attractive for under-resolved turbulence simulations, e.g. [16]. Second, we mimic severe under-resolution of the baseflow by increasing the frequency of the baseflow (40) from  $\pi$  to  $4\pi$ . In combination with the very low number of elements,



**Fig. 8** Spectra of spatial operator, the surface term uses entropy-dissipative Rusanov-type flux (24), the volume term is in skew-symmetric form ( $\alpha = 2/3$ ). Left: Discretisation with polynomial degree  $N = 3$  and 10 elements and linearisation with baseflow shown in Fig. 1. Maximum real part is  $-1.06 \times 10^{-7}$ . Right: Discretisation with polynomial degree  $N = 15$  and 3 elements, and an increased baseflow frequency from  $\pi$  to  $4\pi$ , to mimic severe under-resolution. Zoom in on the spectrum, to show the eigenvalues with maximum real parts of 1.359

this results in large jumps across the element interfaces. A zoom in of the resulting spectrum is plotted in the left part of Fig. 6, where the maximum real part of the eigenvalue is 1.359, which means that the entropy-dissipative scheme in this setup is again locally energy unstable and allows for exponential growth.

As an intermediate conclusion, the numerical results confirm the theoretical findings and the faulty behaviour of the entropy-conserving numerical flux and the entropic split-form approach with diagonal-norm summation-by-parts operators. The schemes allow for a nonphysical exponential growth of local fluctuations. Adding a dissipative mechanism through the numerical fluxes at the interfaces does not necessarily pull the spectrum to the negative half-plane. Neither does necessarily adding dissipation to the volume terms. Both these “fixes”, at best only mask the inherent issue. We suspect that sufficient additional artificial dissipation in the volume terms in combination with truly entropy-dissipative surface fluxes might produce locally correct behaviour. However, that comes at the cost of being over-dissipative and reducing the benefit of a high-order, low dissipation scheme.

On the other hand, the central approximation in these particular investigations is locally energy stable. However, the central approximation of course lacks a non-linear (entropy) stability estimate and hence is not suitable for the simulation of highly non-linear (under-resolved) problems.

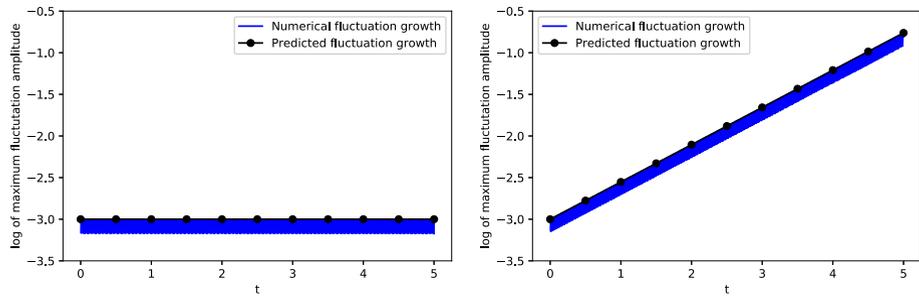
### 2.4.2 Simulation of the Growth of the Fluctuations

The goal of this section is to solidify the numerical results from the previous section and perform actual simulations to monitor the growth of the fluctuations. The semi-discretisation (37) of the non-linear Burgers equation gives a system of weakly coupled ordinary differential equations in time that we integrate with a three-stage third-order strong stability preserving Runge-Kutta scheme with a time step computed as

$$\Delta t = \text{CFL} \frac{h}{(N + 1) \lambda_{\max}}, \tag{42}$$

where  $\lambda_{\max} = \max_u f'(u)$  and CFL = 0.05 to keep the time integration errors low.

We particularly focus on validating the previous results and are hence interested in the evolution of the fluctuations. It is clear that the function shown in Fig. 1 is not a solution of the



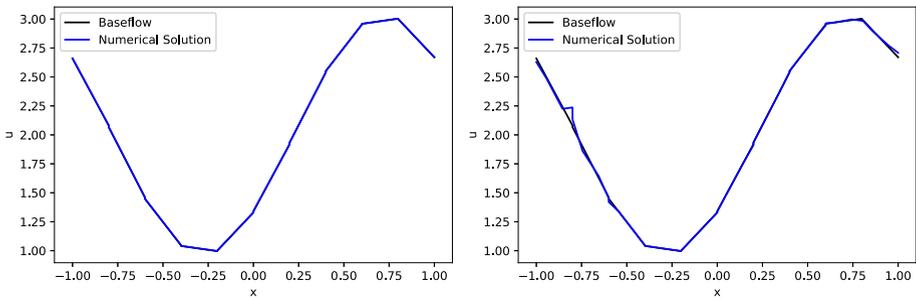
**Fig. 9** Evolution of the maximum amplitude of the fluctuations of the inhomogeneous simulation. Left: Central approximation, volume term in divergence form ( $\alpha = 1$ ) and central numerical flux for the surface term. Right: Entropy-conserving scheme with the skew-symmetric volume terms ( $\alpha = 2/3$ ) and the entropy-conserving flux (18). [Discretisation with 10 elements and polynomial degree  $N = 3$ , using baseflow from Fig. 1 for initialisation and rhs( $\tilde{u}$ ), and an added initial fluctuation being the scaled eigenmode shown in Fig. 3 with the maximum initial amplitude  $10^{-3}$ . End time of the simulation is  $T = 5$ . For comparison, we take the maximum real part of the corresponding spectra and use it to predict the exponential growth. The maximum real part is  $8.8 \times 10^{-8}$  (left) and 1.0307 (right). Spectra shown in Fig. 2]

non-linear homogeneous Burgers equation. Instead, we consider the inhomogeneous Burgers equation, where we use the residual computed with the linearisation state as a discrete source term. For the inhomogenous source term, we first compute the fully non-linear semi-discrete residual with the linearisation state  $\text{rhs}(\tilde{u})$  as a pre-processing step. Then, during the actual simulation, we subtract in every single Runge-Kutta stage the residual of the linearisation state. It is important to note, that we project the linearisation state to piecewise linear polynomials, to guarantee that for a DGSEM approximation with  $N = 3$  or higher, the volume terms in the residual computations are equal for all split-forms including the divergence form and the skew-symmetric form. By solving the inhomogenous Burgers equation with this discrete source term, we make sure that a simulation with initial condition equal to the linearisation state would give zero growth down to machine precision. This is important, as we are now able to add fluctuations  $u'_0(x)$  in a controlled way around the linearisation state to get our actual initial conditions

$$u(x, t = 0) = \tilde{u}(x) + u'_0(x), \tag{43}$$

such that we can monitor the growth of the fluctuations in comparison to the linearisation state.

To bring the simulation results in context with the investigation of the spectra from the previous subsection, we choose the critical eigenmode corresponding to the eigenvalue with the largest real part shown in Fig. 3 as our initial fluctuation  $u'_0(x)$ . Note that the eigenmode is scaled such that its maximum peak is  $10^{-3}$ , thus we have an initial perturbation with amplitude  $10^{-3}$ . Throughout the simulation, we monitor the maximum peak of the approximative solution in comparison to the linearization state  $\tilde{u}$  after every time step and plot the result in log-scale over time in Fig. 9. To compare these numerical results to the previous analysis, we take the maximum real part of the spectra and use this to estimate the growth rate and to generate a prediction of the corresponding exponential growth, starting with the perturbation amplitude  $10^{-3}$ . The spectrum for the central approximation shows indeed no exponential growth, as the amplitude of the fluctuations stay at about the initial amplitude  $10^{-3}$ . In stark contrast, the entropy-conserving discretisation shows significant exponential growth, with a growth rate matching the prediction from the previous analysis of the spectrum.



**Fig. 10** Numerical solution of the inhomogeneous simulation at  $T = 5$  and comparison to baseflow. Left: Central approximation with divergence form volume term ( $\alpha = 1$ ) and the central numerical flux for the surface term. Right: Entropy-conserving scheme with the skew-symmetric volume terms ( $\alpha = 2/3$ ) and the entropy-conserving flux (18). [Discretisation with 10 elements and polynomial degree  $N = 3$ , using baseflow from Fig. 1 for initialisation and rhs( $\tilde{u}$ ), and an added initial fluctuation being the scaled eigenmode shown in Fig. 3 with the maximum initial amplitude  $10^{-3}$ . Spectra shown in Fig. 2]

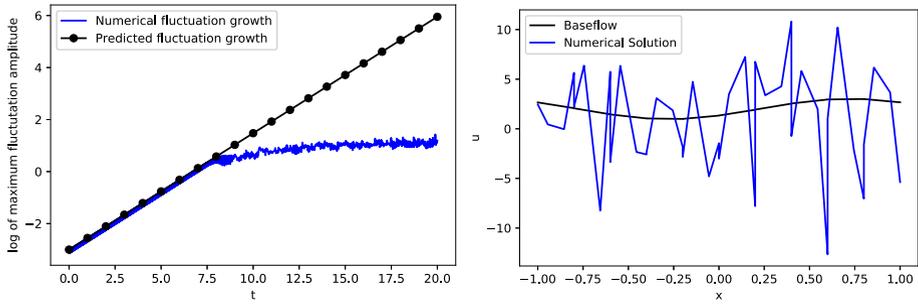
For completeness, we plot the resulting numerical solutions in Fig. 10, where we can clearly observe how the faulty exponential growth of the small scale fluctuations start to get visible in case of the entropy-conserving scheme. The amplitude  $10^{-3}$  of the initial fluctuations grows up to more than 0.1, such that this can already be seen in the simulation results. It is interesting to note again, that the eigenmode is active in the part of the domain with negative slopes, hence the solution first starts to deviate from the baseflow in this part of the domain.

**Remark 8** We did several investigations with different versions of the scheme, different configurations of volume and surface terms, and different initial fluctuation distributions. In every case, a faulty spectrum with a positive maximum real part leads to an exponential growth in the actual simulation and shows that the scheme is locally energy unstable. For discretisations where the maximum real part is zero (in relation to the accuracy of the approximative Jacobian), all simulations keep the initial condition with only negligible variations for all times and are locally energy stable.

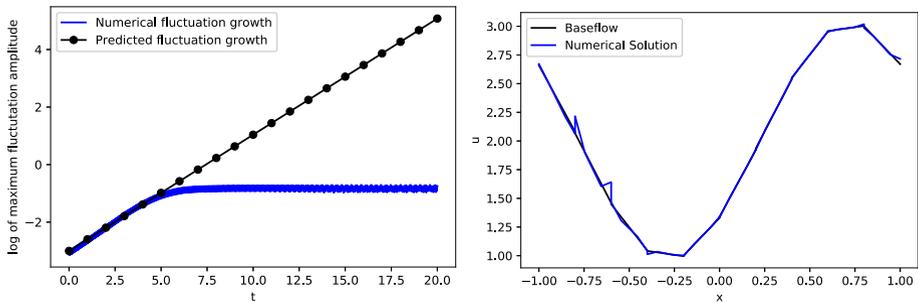
Next, we consider a long term simulation of the entropy-conserving discretisation. The goal is to investigate the behaviour of the method when the amplitude of the fluctuations are so large, that non-linear effects cannot be neglected. We thus increase the end time from  $T = 5$  to  $T = 20$  and show the results in Fig. 11. It is interesting to see that at about  $t = 7.5$ , non-linear behaviour seem to emerge and start to dominate the scheme. It is important to note that the entropy-conserving scheme is neutrally stable in the  $L^2$ -norm, which is clearly visible in the left plot. The right plot shows the numerical solution at the final time, and reveals a wild collection of large scale oscillations that completely dominate the solution and overshadow the baseflow. It is important to recall, the the physical prediction is that the initial condition, the baseflow (black line), stays steady for all times!

We further show an interesting result for the dissipative modification of Tadmor (2.1) and the volume modification of Carpenter et al. discussed in Remark 7. As shown in the previous section, the scheme is not locally energy stable, despite the added dissipation. The left plot of Fig. 12 shows this behaviour with a similar faulty exponential growth, until the non-linear effects kick in at about  $t = 5$  and stabilise the solution via its global  $L^2$  estimate.

Due to the added dissipation, the global stability mechanism kicks in at an earlier time  $t = 5$  compared to the entropy-conserving scheme at time  $t = 7.5$  in Fig. 11. Consequently,



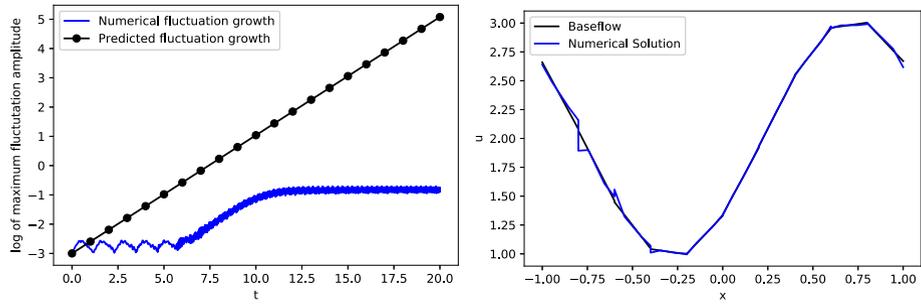
**Fig. 11** Long time behaviour of the inhomogeneous simulation, using the entropy-conserving scheme with skew-symmetric volume terms ( $\alpha = 2/3$ ) and the entropy-conserving flux (18). Left: Evolution of the maximum fluctuation amplitude over time, compared to the predicted exponential growth with the maximum real part 1.0307. Right: Numerical solution at final time  $T = 20$  and comparison with the initial baseflow. [Discretisation with 10 elements and polynomial degree  $N = 3$ , using baseflow from Fig. 1 for initialisation and rhs( $\tilde{u}$ ), and an added initial fluctuation being the scaled eigenmode shown in Fig. 3 with the maximum initial amplitude  $10^{-3}$ ]



**Fig. 12** Long time behaviour of the inhomogeneous simulation, using the modified entropy-dissipative scheme of Tadmor (23) and the volume modification of Carpenter et al. discussed in remark 7. Left: Evolution of the maximum fluctuation amplitude over time, compared to the predicted exponential growth with the maximum real part 0.929. Right: Numerical solution at final time  $T = 20$  and comparison with the initial baseflow. [Discretisation with 10 elements and polynomial degree  $N = 3$ , using baseflow from Fig. 1 for initialisation and rhs( $\tilde{u}$ ), and an added initial fluctuation being the scaled eigenmode shown in Fig. 3 with the maximum initial amplitude  $10^{-3}$ ]

the maximum fluctuation amplitude is smaller in comparison as well and the numerical solution plotted in the left part of Fig. 12 looks much 'nicer' compared to the result of the entropy-conserving scheme in Fig. 11. Still, although it is better than the entropy-conserving result, this should not deflect the attention from the fact that it also displays faulty behaviour.

Finally, we document another interesting numerical result, where the initial fluctuations are not chosen as the particular eigenmode corresponding to the critical eigenvalue with large maximum real part. Instead, we choose a smooth distribution of the initial fluctuation  $u'_0(x) = 0.001 \cos(\pi x)$ . This fluctuation is well resolved by the chosen discretisation. The purpose of this particular setup is to investigate the behaviour of the scheme for smooth initial data under grid refinements, to further investigate if the schemes might be Lax stable. We choose again the scheme with the entropy-dissipative surface modification of Tadmor and volume modification of Carpenter et al. for this assessment. It can be seen in the left



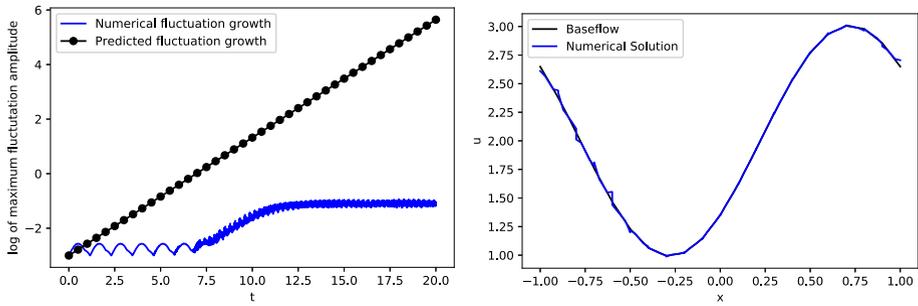
**Fig. 13** Long time behaviour of the inhomogeneous simulation, with a smooth initial fluctuation, using the modified entropy-dissipative scheme of Tadmor (23) and the volume modification of Carpenter et al. discussed in remark 7. Left: Evolution of the maximum fluctuation amplitude over time, compared to the predicted exponential growth with the maximum real part 0.929. Right: Numerical solution at final time  $T = 20$  and comparison with the initial baseflow. [Discretisation with 10 elements and polynomial degree  $N = 3$ , using baseflow from Fig. 1 for initialisation and rhs( $\tilde{u}$ ), and an added smooth initial fluctuation  $u'_0(x) = 0.001 \cos(\pi x)$ ]

plot of Fig. 13 that until about  $t = 5$ , the solution behaves very accurately and keeps the amplitude of the initial fluctuations in check at about  $10^{-3}$ . If one would choose the final time of the simulation smaller than  $t = 5$ , one might get tricked into believing that the scheme works very well. However, the smooth distribution of the fluctuations also contain parts of the eigenmodes that correspond to eigenvalues with positive maximum real part. It comes as no surprise that these eigenmodes grow throughout the simulation and get amplified until they finally dominate the initial smooth fluctuations. Once, the eigenmode dominates, we can observe the exponential growth with the rate 0.9297 until the non-linear behaviour takes over.

If we increase the resolution from 10 to 20 elements, we discussed in the previous section, compare Fig. 7, that the faulty eigenmodes get shifted to higher frequencies. Thus, we expect that for smooth (well-resolved) fluctuations the amplitude of the faulty eigenmodes decrease, as higher frequencies parts containing these modes should decay quickly. As can be seen in Fig. 14, we can observe that indeed the approximative solution stays stable for longer times, up to  $t = 7.5$  in comparison to  $t \approx 6$ . Consequently, the overall maximum growth of the fluctuation amplitude is smaller compared to the lower resolution before the non-linear stability estimate kicks in. However, we can also clearly observe that increasing the resolution only shifts the problem to later times and does not fully resolve it, as we still get artificial exponential growth of the fluctuations at later simulation times. However, these results further strengthen the suggestion, that the schemes might be Lax stable and that they might be convergent for fine enough grids.

The results of the longterm simulations of the entropy-conserving or entropy-dissipative discretisations shown in Figs. 11, 12, 13 and 14 are highly worrisome. They demonstrate that the discretisations might generate results that appear stable, but the simulations are nonphysical as they allow for an exponential growth of small scale fluctuations until the non-linear stability estimate kicks in and provides a global bound. We note that the physical behaviour of the linearised problem with the given positive baseflow is known and does not allow exponential growth.

The worrisome aspect is that the scheme artificially generates solution structures, that are caused by the lack of local energy stability of the numerical scheme and not by actual



**Fig. 14** Long time behaviour of the inhomogeneous simulation, with a smooth initial fluctuation and 20 elements, using the modified entropy-dissipative scheme of Tadmor (23) and the volume modification of Carpenter et al. discussed in remark 7. Left: Evolution of the maximum fluctuation amplitude over time, compared to the predicted exponential growth with the maximum real part 0.929. Right: Numerical solution at final time  $T = 20$  and comparison with the initial baseflow. [Discretisation with 20 elements and polynomial degree  $N = 3$ , using baseflow from Fig. 1 for initialisation and rhs( $\bar{u}$ ), and an added smooth initial fluctuation  $u'_0(x) = 0.001 \cos(\pi x)$ ]

physics. We have also shown that grid refinement might shift the issues to later simulation times and that the schemes might converge for smooth solutions, again, hinting towards Lax stability. At first, one might lean towards rating this behaviour as acceptable in practice, as one could just refine the grid until the simulation time can be reached safely. However, at second sight, this behaviour is at least as worrisome as we do not want to artificially increase the resolution of our discretisation, just to make it stable. This is certainly not a feasible strategy for three-dimensional turbulence. Furthermore, while we can easily assess the behaviour and judge the numerical results for the Burgers equation, simulation results for fluid dynamics equations in multiple spatial dimensions are much more complicated and harder to assess. It seems to be very difficult to decide, when the resolution in a complex application is enough to suppress the artificial exponential growth (assuming that it is even feasible to increase the resolution). In particular, one may then ask the critical question, if those artificial structures generated by the scheme might get transformed by the high-order scheme into something that resembles a meaningful fluid dynamics structure. One might even interpret the extended range of medium and small scales appearing as a positive aspect of the “low numerical diffusion of the high-order approximation”.

To further investigate this hypothesis, we have to consider the true flow equations, which is the topic of the next section.

### 3 Local Stability Analysis for the Compressible Euler Equations

In this section, we show that the issues of the entropy-conserving fluxes carry over to the compressible Euler equations. To simplify the analysis, we first consider the one-dimensional version

$$u_t + f(u)_x = \begin{pmatrix} \rho \\ \rho v \\ \rho E \end{pmatrix}_t + \begin{pmatrix} \rho v \\ \rho v^2 + p \\ \rho E v + p v \end{pmatrix}_x = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{44}$$

where  $\rho$  is the density,  $v$  the velocity,  $E$  the specific total energy,  $p = (\gamma - 1) \rho (E - v^2/2)$  the pressure with the perfect gas assumption, and  $\gamma = 1.4$  the adiabatic coefficient. For our analysis, we consider the entropy pair

$$U(u) = -\frac{\rho s}{\gamma - 1}, \quad F(u) = -\frac{\rho v s}{\gamma - 1}, \quad s(u) = \ln\left(\frac{p}{\rho^\gamma}\right) \tag{45}$$

where  $s$  is the thermodynamic entropy. The corresponding entropy variables are

$$w = \frac{\partial U}{\partial u} = \left(\frac{\gamma - s}{\gamma - 1} - \frac{\rho}{p} \frac{v^2}{2}, \frac{\rho}{p} v, -\frac{\rho}{p}\right)^T. \tag{46}$$

We consider first the finite volume discretisation (3) of the compressible Euler Eq. (44). There are several numerical flux functions available in literature, e.g., [4,21,35] that give discrete entropy-conservation. We consider for instance the variant presented by Chandrashekar [4] that is given by

$$f_{i+1/2}^{EC} = \begin{pmatrix} \{\rho\}^{\ln} \{\{v\}\} \\ \{\rho\}^{\ln} \{\{v\}\}^2 + \hat{p} \\ \{\rho\}^{\ln} \{\{v\}\} \hat{h} \end{pmatrix}, \tag{47}$$

where  $\{\cdot\}$  is the arithmetic mean value, the logarithmic mean is

$$\{\rho\}^{\ln} = \frac{\rho_{i+1} - \rho_i}{\ln(\rho_{i+1}) - \ln(\rho_i)}, \tag{48}$$

and

$$\hat{p} = \frac{\{\rho\}}{\{\rho/p\}}, \quad \hat{h} = \frac{1}{\{\rho/p\} \ln(\gamma - 1)} + \{\{v\}\}^2 - \frac{1}{2} \{\{v^2\}\} + \frac{\hat{p}}{\{\rho\}^{\ln}}. \tag{49}$$

We consider a particular simple test problem with periodic boundary conditions in the domain  $[-1, 1]$ , where the initial condition is given by a density wave

$$\rho(x) = 1 + 0.98 \sin(2\pi x), \quad v = 0.1, \quad p = 20, \tag{50}$$

with a minimum density of  $\rho_{\min} = 0.02$ . Plugging this solution into the compressible Euler Eq. (44), the problem reduces to a set of constant coefficient advection equations. Thus, the exact solution of this problem is directly given by the characteristics theory, i.e. the wave is transported without changing its shape or form. As this test problem reduces the compressible Euler equations to a simple advection problem, we know that again the finite volume scheme with the central numerical flux

$$f_{i+1/2}^{CN} = \frac{1}{2}(f(u_i) + f(u_{i+1})) \tag{51}$$

is the least diffusive scheme that is energy stable. Moreover, all three equations reduce to exactly the same advection equation when using the central flux which is well-known to produce accurate solutions. Thus, we are again able to use the central flux as the baseline scheme for comparisons with other numerical flux functions.

We turn to the analysis of the mass equation and compare the entropy-conserving flux (47) to the central flux (51) and get the relationship

$$(f_1)_{i+1/2}^{EC} = (f_1)_{i+1/2}^{CN} - \frac{1}{2} (R_1)_{i+1/2}^{EC} (\rho_{i+1} - \rho_i), \tag{52}$$

where the dissipation coefficient is

$$(R_1)_{i+1/2}^{\text{EC}} = \frac{2(\{\rho v\} - \{\rho\}^{\text{ln}} \{v\})}{\rho_{i+1} - \rho_i} = (\{\rho\} - \{\rho\}^{\text{ln}}) \frac{2\{v\}}{\rho_{i+1} - \rho_i} + \frac{(v_{i+1} - v_i)}{2}, \quad (53)$$

which shows that again anti-dissipation could occur for certain configurations, as  $(R_1)_{i+1/2}^{\text{EC}}$  is indefinite. In particular, if we consider the test problem (50) with the constant positive velocity, and since  $\{\rho\}^{\text{ln}} \leq \{\rho\}$ , the dissipation coefficient  $(R_1)_{i+1/2}^{\text{EC}}$  is negative in the part of the domain where the density slope is negative, i.e.  $\rho_x < 0$ .

**Remark 9** Unlike the central scheme, the entropy-conservative flux does not reduce to the same scheme for the advection equation for all three equations. Hence, the numerical solutions of the three equations will not be the same but interact with each other. Consequently, the scheme does not only introduce negative artificial dissipation, but also contains a mechanism for generating perturbations.

This short analysis for the simple test case already demonstrates that the issues of anti-dissipation carry over from the discussion on the Burgers equation to the compressible Euler equations. We note that the entropy-conserving numerical flux is again the main building block to extend the finite volume approach to high-order discretisations [2,9,15]. Thus, it is expected that the issues with anti-dissipation and lack of local energy stability most likely carry over to the high-order framework.

In the following numerical results, we do not simulate the evolution of the fluctuations with respect to a baseflow by using an inhomogeneous test case. Instead, we perform a standard simulation with the non-linear homogeneous compressible Euler equations. We are interested in numerical investigations of the actual behaviour of the high-order framework. We choose again a high-order discontinuous Galerkin spectral element method with Legendre-Gauss-Lobatto nodes, which belongs to the class of diagonal-norm summation-by-parts methods. In contrast to the Burgers equation, no explicit split-form to obtain entropy-conservation is known up to this point. Instead, Fisher [9] and Carpenter et al. [2,3] extended the high-order entropy-conserving reconstruction of LeFloch et al. [25] to diagonal-norm SBP operators on finite domains, including e.g. the spectral collocation operator with Legendre-Gauss-Lobatto nodes. For a detailed description of the algorithms in multiple dimensions and additional assessment of their properties, we further refer to [15,17].

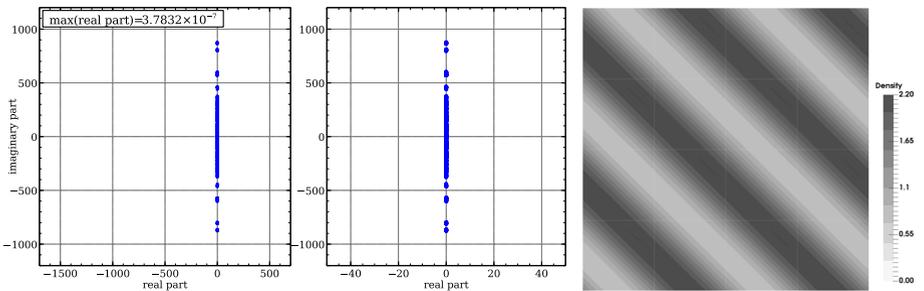
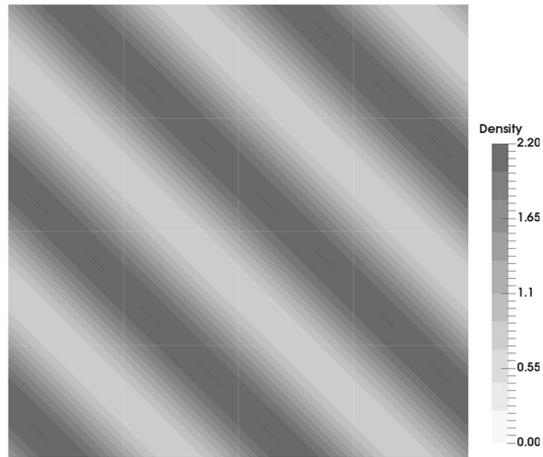
The numerical investigations are carried out with the open source three-dimensional high-order simulation code Fluxo ([github.com/project-fluxo](https://github.com/project-fluxo)) and for cross-evaluation with the open source Julia language based simulation framework Trixi.jl ([github.com/trixi-framework/Trixi.jl](https://github.com/trixi-framework/Trixi.jl)). The assessment is done with  $4 \times 4$  elements in 2D in the domain  $[-1, 1]^2$  with polynomial degree  $N = 5$ . The initial condition (50) is extended to multiple spatial dimension and reads

$$\rho(x) = 1 + A \sin(2\pi(x + y)), \quad A = 0.98, \quad (v_1, v_2, v_3) = (0.1, 0.2, 0), \quad p = 20. \quad (54)$$

Figure 15 shows the visualisation of the initialised density. We observe that this particular test problem is very well resolved. Furthermore, this initial distribution is also used as a linearisation state to compute the spectrum of the non-linear operators.

The semi-discretisation is explicitly integrated in time with a low-storage 5-stage Runge-Kutta method of 4th order accuracy [22]. The time step is updated during the simulation with  $\text{CFL} = 0.05$ . The end time of the simulation is  $T = 5$ , but we will add the final crash times for simulations that did not reach the end time. For the investigation of the spectra,

**Fig. 15** Visualisation of the discrete density distribution of the initial condition (54) used for the numerical simulations, on a  $4 \times 4$  grid with polynomial degree  $N = 5$

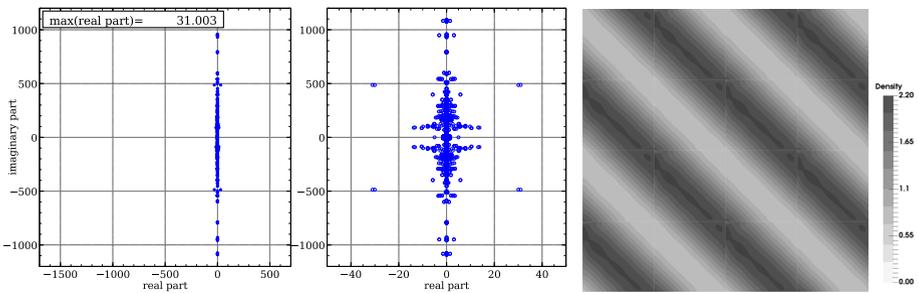


**Fig. 16** Numerical results with the central approximation, volume term in divergence form and the central numerical flux function for the surface term. Left: Spectrum of the spatial operator at  $t = 0$ , maximum real part is  $3.783 \times 10^{-7}$ . Middle: Zoom in view of the spectrum. Right: Visualisation of the density distribution at the final simulation time  $T = 5$ . [The grid is  $4 \times 4$  with polynomial degree  $N = 5$ . The initial condition is shown in Fig. 15]

we use the initial conditions as our linearisation state  $\tilde{u}(x, y, z) = u_0(x, y, z)$ . The spectra are generated with Trixi.jl, which features automatic differentiation. Thus, the Jacobians are computed exactly, the in-built linear algebra package of Julia is then used to compute the eigenspectrum.

As a first result, we verify in Fig. 16 the behaviour of the central approximation, where we use the volume terms in divergence form and the numerical surface flux as the central flux. As can be seen in the left plot, the spectrum is essentially imaginary, with the largest positive real part being  $3.783 \times 10^{-7}$ . To strengthen this assumption we performed a long time simulation with the central approximation up to  $T = 200$ , which translates to over 1.6 million RK time steps without observing any spurious growth at all.

The right plot shows the corresponding simulation results at the final time  $T = 5$ , and confirms that the solution is well resolved and properly advected by the dissipation-free purely central high-order approximation.



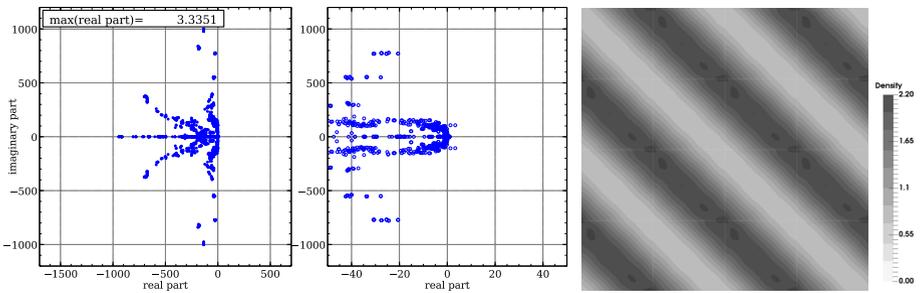
**Fig. 17** Numerical results with the entropy-conserving approximation, using entropic flux-difference volume terms [2] with the entropy-conserving flux of Chandrashekar [4], which is also used for the surface term. Left: Spectrum of the spatial operator at  $t = 0$ , maximum real part is 31.003. Middle: Zoom in view of the spectrum. Right: Visualisation of the density at time  $t = 0.55$ , right before the code crashes because of negative density. [The grid is  $4 \times 4$  with polynomial degree  $N = 5$ . The initial condition is shown in Fig. 15]

Next, we use the entropy-conserving high-order approximation based on the numerical flux function by Chandrashekar [4]. (We only document results with this flux, but we also tested other entropy-conservative fluxes including the one by Ismail and Roe [21] with similar outcome). Figure 17 shows the results of this investigation. The left plot shows the spectrum of the operator and reveals that there are significant positive real parts with a maximum value of about 31.003, which confirms our theoretical analysis from above. For strong negative density gradients, the scheme produces anti-dissipation which in our terminology implies a local energy instability. The impact on the simulation is drastic for this particular test problem, as the simulation crashes at time  $t = 0.5533$  before reaching the final time  $T = 5$ . We note that for illustration, the test case is setup on purpose such that severe faulty behaviour in the discrete density evolution can cause negative density. However, we stress again that the simple central approximation is able to comfortably reach the final time  $T = 5$ , and much longer end times without problems. This is a simple well resolved test problem and it is thus even more worrisome that the entropy-conserving high-order discretisation fails. Furthermore, we also tested with other setups with less density variation and we could observe similar issues, it just requires longer simulation times for the instability to grow and crash the simulation.<sup>1</sup> We emphasise that the behaviour is independent of the choice of CFL, as it is not a time integration issue but a spatial discretisation failure.

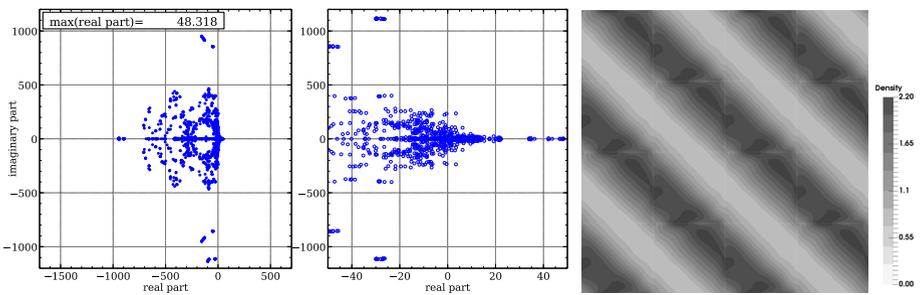
We next switch on the surface dissipation and use the Rusanov flux [45]. The high-order scheme is now formally entropy-dissipative. We note again that the volume terms are still in entropy-conserving form and it is not obvious that a surface dissipation term can fully control anti-dissipation generation through the volume. The results are plotted in Fig. 18, which shows that the maximum real part is 3.3351 and hence the discretisation is locally energy unstable again. Indeed, the simulation crashes at an early time  $t = 0.6595$ .

While we focus on the discretisations with entropic properties, it is interesting to note that there are many different split-forms for compressible fluid dynamics, with different properties such as e.g. kinetic-energy-preservation, see e.g. [5,7,15,23,31]. These split-forms

<sup>1</sup> i.e. for a density variation  $\rho = 1 + 0.5 \sin(2\pi(x + y))$ , the entropy-conserving scheme crashes at  $T = 26.7$ .



**Fig. 18** Numerical results with the entropy-dissipative approximation, using entropic flux-difference volume terms [2] with the entropy-conserving flux of Chandrashekar [4], and guaranteed entropy-dissipative surface terms via the Rusanov flux [45]. Left: Spectrum of the spatial operator at  $t = 0$ , maximum real part is 3.3351. Middle: Zoom in view of the spectrum. Right: Visualisation of the density at time  $t = 0.65$ , right before the code crashes because of negative density. [The grid is  $4 \times 4$  with polynomial degree  $N = 5$ . The initial condition is shown in Fig. 15]



**Fig. 19** Numerical results with the kinetic-energy-dissipative approximation, using split-form volume terms [15] based on the Kennedy and Gruber splitting [23] and surface terms with the dissipative Rusanov flux [45]. Left: Spectrum of the spatial operator at  $t = 0$ , maximum real part is about about 48.318. Middle: Zoom in view of the spectrum. Right: Visualisation of the density at time  $t = 0.05$ , right before the code crashes with negative density. [The grid is  $4 \times 4$  with polynomial degree  $N = 5$ . The initial condition is shown in Fig. 15]

are all different, but have one obvious common property: they are not the central approximation. Consequently, all these split-forms can be compared to the central approximation and have additional terms, that are either dissipative or maybe anti-dissipative. Hence, if for our considered baseflow the split-form does not reduce to the central approximation, the split-form has a similar issue with local energy stability as the entropy-conserving variant. As an example and numerical evidence, we consider the Kennedy and Gruber splitting [23] with the Rusanov numerical flux [45] for the surface terms. This gives a discretisation that is kinetic-energy-dissipative [15]. The numerical results are shown in Fig. 19 and clearly show faulty behaviour for this test problem. The maximum real part is about 48.318, which again reveals that the discretisation is locally energy unstable. Again the impact on the actual simulation is strong, as the simulation crashes at an early time  $t = 0.0845$ .

In “Appendix” A, we summarise additional simulation results obtained for the same density wave test case and the same schemes presented in this section. In particular, we compare the time evolution of the  $L_2$  error in density for the different schemes, on a  $4 \times 4$  and a  $8 \times 8$  grid and also for two ‘weaker’ amplitudes  $A$  of the initial density from (54). The results

underline that the undesired behaviour of the EC and KG split-form schemes is not limited to one particular setup of the density wave and while being reduced (another hint that these schemes are Lax stable) does not disappear completely on the finer grid or when decreasing the density gradients.

## 4 Conclusion

In this work, we demonstrate worrisome and sometimes downright faulty behaviour of the recently developed class of entropic high-order summation-by-parts schemes (including split-forms). To articulate the issue, we introduced the notion of local energy stability, i.e. the stability of fluctuations around a linearisation state for the non-linear problem. The core of the issue is that discretisation of the non-linear PDE and linearisation of the resulting scheme is not the same as discretising the linearised equations. Hence, non-linear stability does not automatically give local energy stability.

To study the stability properties of the numerical schemes, we deliberately choose test cases, where the continuous problem has no exponential growth. For these special test cases, we could use the central approximation based on the simple central flux as a baseline discretisation to obtain the least diffusive scheme that is locally energy stable. We compare entropy-conserving fluxes to the central flux and observe that the difference can either be dissipative or anti-dissipative. Artificial anti-dissipation is problematic, as it tries to artificially “push” the solution away from the linearisation state, i.e. it generates artificial exponential growth of small scale fluctuations. The theoretical findings have been verified in numerical tests, where the corresponding spectra of the spatial operator show significant positive real parts for continuous problems with no exponential growth and the simulations demonstrate actual artificial exponential growth of fluctuations. We observed that adding a dissipative mechanism through the numerical fluxes at the interfaces does not necessarily guarantee local energy stability either, and neither does adding dissipation to the volume terms. Both these “fixes” at best only mask the inherent issue.

We further show that the faulty behaviour carries over to the compressible Euler equations with a potential negative impact on the robustness of these schemes. Again, we choose a smooth exact solution of the compressible Euler equations that has no exponential growth. While the simple central approximation, without added dissipation, happily runs this test case, the entropy-conserving and the entropy-dissipative schemes crash at early times due to nonphysical exponential growth of density fluctuations. However, the central approximation lacks a non-linear stability estimate and thus cannot be directly applied for strongly non-linear problems, such as e.g. under-resolved turbulence. As a final remark, we demonstrate that this faulty behaviour also appears in other split-forms used in the fluid dynamics community, including kinetic-energy-preserving variants such as the Kennedy and Gruber splitting, with similar worrisome numerical results as shown in our numerical investigations. As a matter of fact, the issue carries over to all split-forms that do not reduce to a central discretisation for the particular test cases considered in this work. However, the central discretisation is not preferable either as it does not have a non-linear global bound.

Our theoretical derivations and numerical investigations helped us to rule out energy stability. However, the derivations also show that the anti-dissipation is proportional to the

mesh size  $h$  and thus the schemes might be Lax stable. We have thus investigated the impact of grid refinement on the spectra and the exponential growth. Grid refinement steps do not change the magnitude of the faulty eigenvalues, but do shift them to higher frequencies. Furthermore, the faulty exponential growths of smooth fluctuations get pushed to later times when refining the grid. These investigations indicate that (i) for a given simulation end time, it is possible to choose a grid resolution where the erroneous exponential growth is not dominating the results yet; (ii) the entropic (including split-forms) schemes might be Lax stable and hence the scheme might be converging for smooth solutions as the mesh size  $h \rightarrow 0$ . For practical simulations with finite grid size  $h > 0$ , however, it is clearly not a desirable behaviour of any high-fidelity scheme, that the grid resolution has to be adjusted solely to push the faulty behaviour to later times such that they do not affect the results.

It is clear that the erroneous behaviour of the high-order entropic split-form schemes needs further investigations and a proper fix. We tested several mechanisms to introduce dissipation through the volume, but it seems that it is not straightforward to find a strategy that does provide enough dissipation for local energy stability without overwhelming the high-order accuracy with excessive dissipation. The goal is to construct a scheme that is locally energy stable, has a non-linear entropy bound and is still high-order accurate. So, far, we are only able to get two of these properties when using the high-order SBP split-form approach. Thus, further research is needed and this work is a first documentation of the issues. Without a fix, results obtained with these schemes should be interpreted carefully. Due to the locally energy unstable behaviour, nonphysical structures might get introduced and subsequently evolved by the fluid equation to appear as 'meaningful' vortices and turbulent-like features, while being purely a numerical artefact.

**Acknowledgements** The authors thank Dr. David Kopriva, Dr. Hendrik Ranocha and Dr. Juan Manzanero for the discussion on linear stability of split-forms, linearisation and non-linear stability. We further thank Dr. Hendrik Ranocha for the extension of `Trixi.jl` with automatic differentiation capabilities that we use for the 2D compressible Euler results.

**Funding** Open Access funding enabled and organized by Projekt DEAL. Gregor Gassner was supported by the European Research Council (ERC) under the European Union's Eight Framework Program Horizon 2020 with the research project Extreme, ERC Grant Agreement No. 714487.

**Code Availability** For the numerical results, prototype simulation software was developed and is available for everyone upon request from the authors if not available as open source software.

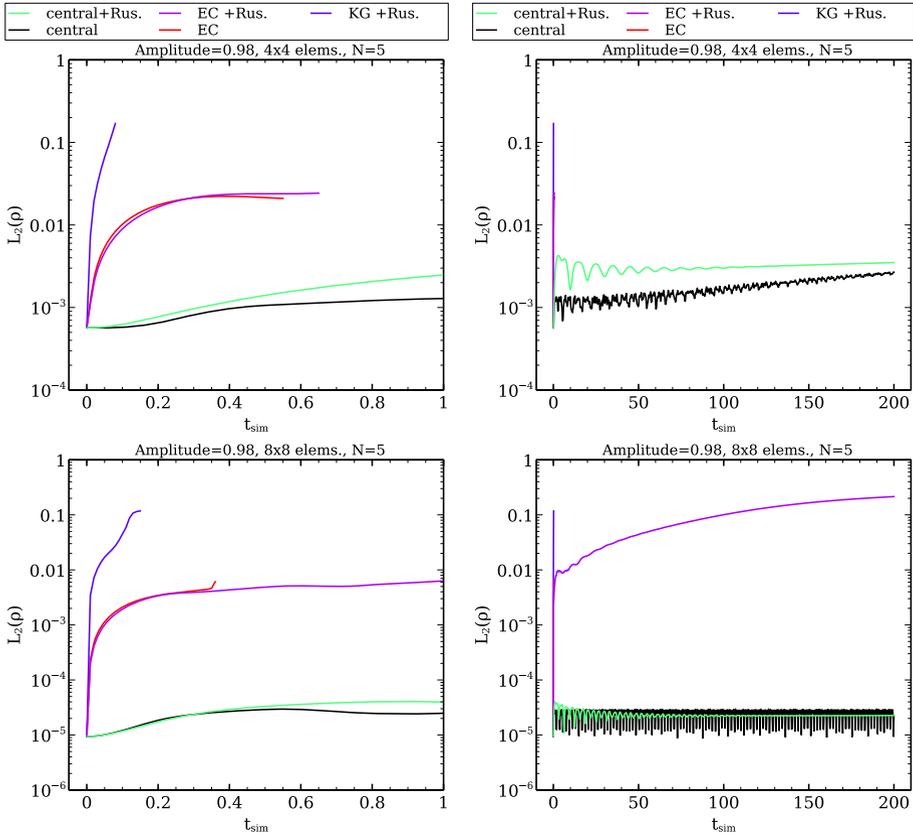
## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Additional Simulation Results for the 2D Density Wave

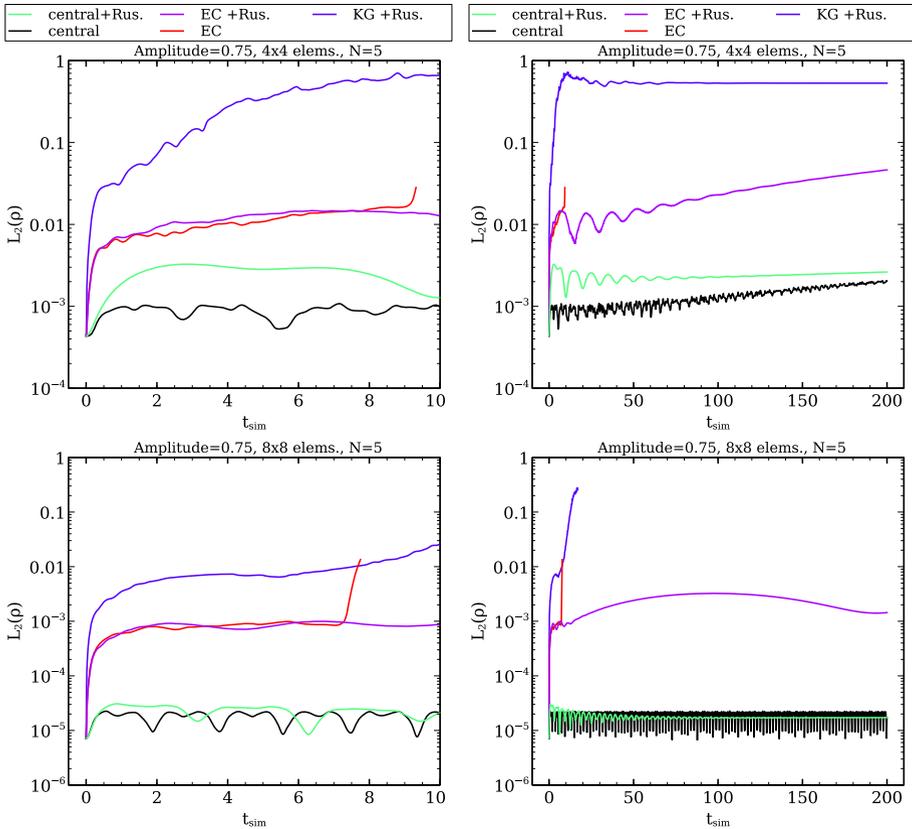
In this section, we summarise additional simulation results obtained for the nonlinear compressible Euler equations without source terms and the *same* two-dimensional density wave (54) of Sect 3, where the density is initialised as a wave with an amplitude  $A = 0.98$  on a  $4 \times 4$  grid and a polynomial degree  $N = 5$ . In particular, since the exact solution is known,



**Fig. 20**  $L_2$  error of density over simulation time, for initial density amplitude  $A = 0.98$  on  $4 \times 4$  grid (top row) and  $8 \times 8$  grid (bottom row). Left the initial phase  $t_{sim} \leq 1$  is shown, and  $t_{sim} \leq 200$  on the right

we compare the time evolution of the  $L_2$  error in density up to  $T = 200$  for all discretisation choices of volume and surface fluxes presented in Sect 3, namely the purely central scheme ('central') and the central scheme with Rusanov surface flux ('central+Rus.'). the entropy and kinetic energy preserving flux by Chandrashekar ('EC'), either with the same flux at the surface or with Rusanov ('EC+Rus.'). and also the split-form Kennedy and Gruber flux with Rusanov surface flux ('KG+Rus.').

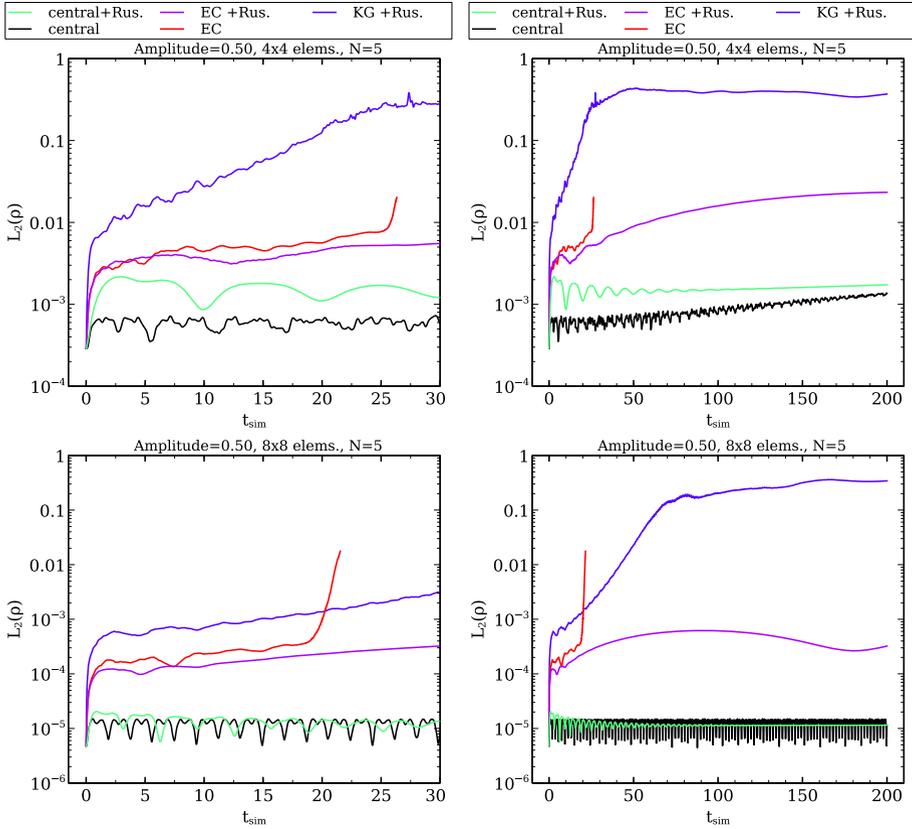
The evolution of the errors for all schemes is shown in Fig. 20, together with the results using a finer grid of  $8 \times 8$ . Additionally, the same analysis is repeated for two 'weaker' density amplitudes  $A = 0.75$ , shown in Fig. 21 and  $A = 0.50$ , shown in Fig. 22.



**Fig. 21**  $L_2$  error of density over simulation time, for initial density amplitude  $A = 0.75$  on  $4 \times 4$  grid (top row) and  $8 \times 8$  grid (bottom row). Left the initial phase  $t_{sim} \leq 10$  is shown, and  $t_{sim} \leq 200$  on the right

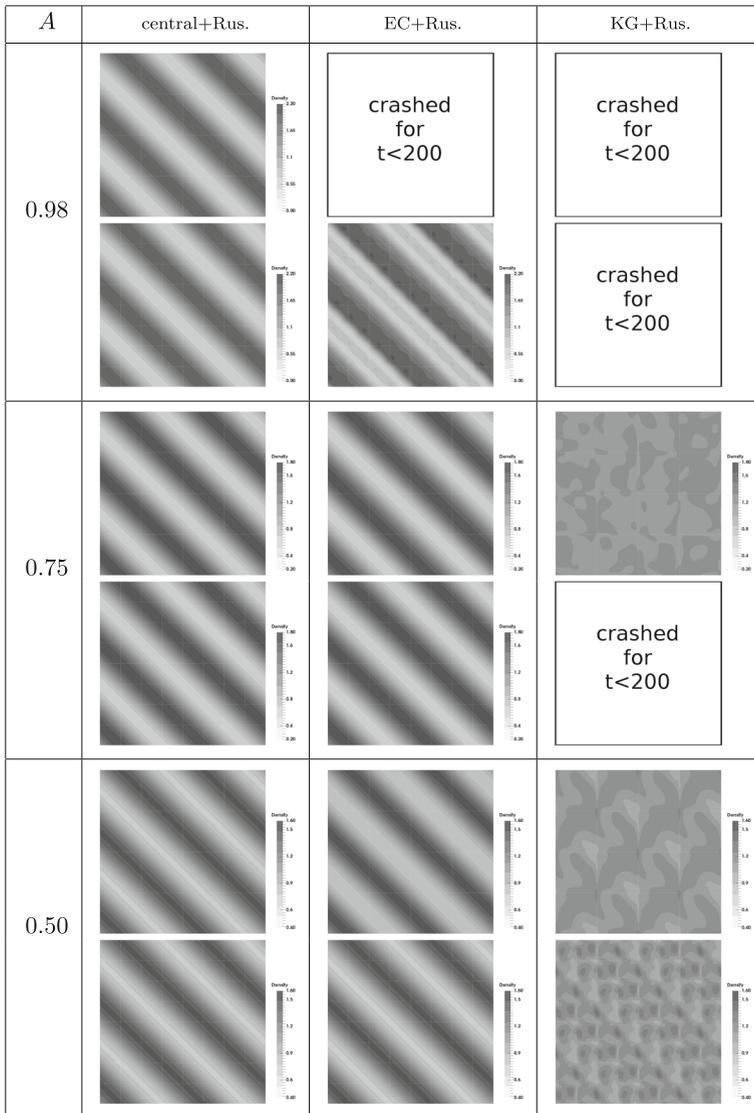
The results clearly show that for the central scheme with a central or a Rusanov surface flux, the error initially changes but remains on the same level until  $T = 200$ . The EC scheme shows a large initial increase of the error and crashes, due to negative density, for all amplitudes  $A$ ; the larger the amplitude, the earlier it crashes. The EC scheme with Rusanov surface flux suffers from the same initial error increase as the EC scheme, but can recover the wave with a much larger error, except in the case of  $A = 0.98$ . This is in perfect agreement with the finding that the anti-dissipation of the EC scheme is related to the density gradient, and the Rusanov dissipation acts at later times to counteract the growth of the solution. Finally, the KG scheme with Rusanov flux has the same issues as the EC scheme and either crashes or the errors become very large.

Regarding the results on the finer  $8 \times 8$  grid, we observe the same behaviour, but on a lower error level. The EC and KG schemes show again a strong initial error increase and either crash or introduce significantly higher error levels.



**Fig. 22**  $L_2$  error of density over simulation time, for initial density amplitude  $A = 0.5$  on  $4 \times 4$  grid (top row) and  $8 \times 8$  grid (bottom row). Left the initial phase  $t_{sim} \leq 30$  is shown, and  $t_{sim} \leq 200$  on the right

For completeness, in Fig. 23, we plot the density distribution at the end of the simulation ( $t = 200$ ), discarding crashed simulations, for the central, EC and KG scheme with Rusanov surface flux, and for both grids and the three choices of the amplitude  $A$ . The large error of the KG scheme is clearly visible, for the EC scheme, the wave still looks intact, but as reported, has a much larger error.



**Fig. 23** Final density distribution at  $T = 200$  for the initial density amplitudes  $A = 0.98/0.75/0.5$ , with a Rusanov surface flux and the volume flux being either central (left), EC (middle) or KG (right), for a  $4 \times 4$  grid (upper row) and  $8 \times 8$  grid (lower row)

## References

1. Barth, T.J.: Numerical methods for gasdynamic systems on unstructured meshes. In: An Introduction to Recent Developments in Theory and Numerics for Conservation Laws, pp. 195–285. Springer, Berlin Heidelberg (1999)
2. Carpenter, M., Fisher, T., Nielsen, E., Frankel, S.: Entropy stable spectral collocation schemes for the Navier–Stokes equations: discontinuous interfaces. *SIAM J. Sci. Comput.* **36**(5), B835–B867 (2014)

3. Carpenter, M.H., Fisher, T., Nielsen, E., Parsani, M., Svård, M., Yamaleev, N.: Entropy stable summation-by-parts formulations for compressible computational fluid dynamics. In R. Abgrall, C. Shu (eds.) *Handbook of Numerical Analysis*, vol. 17, chapter 10, pp. 495–524. Elsevier, Amsterdam (2016)
4. Chandrashekar, P.: Kinetic energy preserving and entropy stable finite volume schemes for compressible Euler and Navier–Stokes equations. *Commun. Comput. Phys.* **14**, 1252–1286 (2013)
5. Coppola, G., Capuno, F., Pirozzoli, S., de Luce, L.: Numerically stable formulations of convective terms for turbulent compressible flows. *J. Comput. Phys.* **382**, 86–104 (2019)
6. Crean, J., Hicken, J.E., Fernández, D.C.D.R., Zingg, D.W., Carpenter, M.H.: Entropy-stable summation-by-parts discretization of the Euler equations on general curved elements. *J. Comput. Phys.* **356**, 410–438 (2018). <https://doi.org/10.1016/j.jcp.2017.12.015>
7. Ducros, F., Laporte, F., Soulères, T., Guinot, V., Moinat, P., Caruelle, B.: High-order fluxes for conservative skew-symmetric-like schemes in structured meshes: application to compressible flows. *J. Comput. Phys.* **161**, 114–139 (2000)
8. Evans, L.: Weak convergence methods for nonlinear partial differential equations. *Am. Math. Soc.* (1990). <https://doi.org/10.1090/cbms/074>
9. Fisher, T.: High-order  $l^2$  stable multi-domain finite difference method for compressible flows. Ph.D. thesis, Purdue University (2012)
10. Fisher, T., Carpenter, M., Nordström, J., Yamaleev, N.K., Swanson, C.: Discretely conservative finite-difference formulations for nonlinear conservation laws in split form: theory and boundary conditions. *J. Comput. Phys.* **234**, 353–375 (2013)
11. Fisher, T.C., Carpenter, M.H.: High-order entropy stable finite difference schemes for nonlinear conservation laws: finite domains. *J. Comput. Phys.* **252**, 518–557 (2013)
12. Fjordholm, U., Mishra, S., Tadmor, E.: Arbitrarily high order accurate entropy stable essentially non-oscillatory schemes for systems of conservation laws. *SIAM J. Numer. Anal.* **50**(2), 544–573 (2012)
13. Friedrich, L., Winters, A.R., Fernández, D.C.D.R., Gassner, G.J., Parsani, M., Carpenter, M.H.: An entropy stable  $h/p$  non-conforming discontinuous Galerkin method with the summation-by-parts property. *J. Sci. Comput.* **77**(2), 689–725 (2018). <https://doi.org/10.1007/s10915-018-0733-7>
14. Gassner, G.: A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to SBP-SAT finite difference methods. *SIAM J. Sci. Comput.* **35**(3), A1233–A1253 (2013)
15. Gassner, G., Winters, A., Kopriva, D.A.: Split form nodal discontinuous Galerkin schemes with summation-by-parts property for the compressible Euler equations. *J. Comput. Phys.* **327**, 39–66 (2016)
16. Gassner, G.J., Beck, A.D.: On the accuracy of high-order discretizations for underresolved turbulence simulations. *Theoret. Comput. Fluid Dyn.* **27**(3–4), 221–237 (2013)
17. Gassner, G.J., Winters, A.R., Hindenlang, F.J., Kopriva, D.A.: The BR1 scheme is stable for the compressible Navier–Stokes equations. *J. Sci. Comput.* **77**(1), 154–200 (2018)
18. Gustafsson, B., Kreiss, H.O., Olliger, J.: *Time Dependent Problems and Difference Methods*. John Wiley and Sons Inc, New Jersey (1995)
19. Honein, A.E., Moin, P.: Higher entropy conservation and numerical stability of compressible turbulence simulations. *J. Comput. Phys.* **201**(2), 531–545 (2004)
20. Ismail, F., Roe, P.: Affordable, entropy-consistent euler flux functions II: Entropy production at shocks. *J. Comp. Phys.* pp. 5410–5436 (2009)
21. Ismail, F., Roe, P.L.: Affordable, entropy-consistent Euler flux functions II: entropy production at shocks. *J. Comput. Phys.* **228**(15), 5410–5436 (2009)
22. Kennedy, C., Carpenter, M.: Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Appl. Numer. Math.* **44**, 139–181 (2003)
23. Kennedy, C.A., Gruber, A.: Reduced aliasing formulations of the convective terms within the Navier–Stokes equations for a compressible fluid. *J. Comput. Phys.* **227**, 1676–1700 (2008)
24. Lax, P., Richtmyer, R.D.: Survey of the stability of linear finite difference equations. *Commun. Pure Appl. Math.* **9**(2), 267–293 (1956)
25. LeFloch, P., Rohde, C.: High-order schemes, entropy inequalities, and nonclassical shocks. *SIAM J. Numer. Anal.* **37**(6), 2023–2060 (2000)
26. LeFloch, P.H., Mercier, J.M., Rohde, C.: Fully discrete, entropy conservative schemes of arbitrary order. *SIAM J. Numer. Anal.* **40**(5), 1968–1992 (2002)
27. LeFloch, P.H., Rohde, C.: High-order schemes, entropy stable inequalities and nonclassical shocks. *SIAM J. Numer. Anal.* **37**(6), 2023–2060 (2000)
28. Madrane, A., Fjordholm, U.S., Mishra, S., Tadmor, E.: Entropy conservative and entropy stable finite volume schemes for multi-dimensional conservation laws on unstructured meshes. In *ECCOMAS 2012* (2012)
29. Mankbadi, R.R.: Linear stability theory. In: *Transition, Turbulence, and Noise*, pp. 21–49. Springer, New York (1994)

30. Manzanero, J., Rubio, G., Ferrer, E., Valero, E., Kopriva, D.A.: Insights on aliasing driven instabilities for advection equations with application to gauss-lobatto discontinuous galerkin methods. *J. Sci. Comput.* **75**(3), 1262–1281 (2017)
31. Morinishi, Y.: Skew-symmetric form of convective terms and fully conservative finite difference schemes for variable density low-mach number flows. *J. Comput. Phys.* **229**(2), 276–300 (2010)
32. Murman, S.M., Diosady, L., Garai, A., Ceze, M.: A space-time discontinuous-Galerkin approach for separated flows. In 54th AIAA Aerospace Sciences Meeting, p. 1059 (2016)
33. Parsani, M., Carpenter, M., Nielsen, E.: Entropy stable discontinuous interfaces coupling for the three-dimensional compressible Navier–Stokes equations. *J. Comput. Phys.* **290**(C), 132–138 (2015)
34. Parsani, M., Carpenter, M.H., Fisher, T.C., Nielsen, E.J.: Entropy stable staggered grid discontinuous spectral collocation methods of any order for the compressible Navier–Stokes equations. *SIAM J. Sci. Comput.* **38**(5), A3129–A3162 (2016). <https://doi.org/10.1137/15m1043510>
35. Ranocha, H.: Comparison of some entropy conservative numerical fluxes for the euler equations. *J. Sci. Comput.* **76**(1), 216–242 (2017)
36. Ray, D., Chandrashekar, P.: An entropy stable finite volume scheme for the two dimensional Navier–Stokes equations on triangular grids. *Appl. Math. Comput.* **314**, 257–286 (2017)
37. Sandham, N.D., Li, Q., Yee, H.C.: Entropy splitting for high-order numerical simulation of compressible turbulence. *J. Comput. Phys.* **178**(2), 307–322 (2002)
38. Sjögreen, B., Yee, H.C., Kotov, D.: Skew-symmetric splitting and stability of high order central schemes. *J. Phys: Conf. Ser.* **837**(1), 012019 (2017)
39. Strang, G.: Accurate partial difference methods II. Non-linear problems. *Numer. Math.* **6**, 37–46 (1964)
40. Svård, M., Lundberg, J., Nordström, J.: A computational study of vortex–airfoil interaction using high-order finite difference methods. *Comput. Fluids* **39**(8), 1267–1274 (2010)
41. Tadmor, E.: The numerical viscosity and the entropy condition for conservative difference schemes. *Math. Comp.* **43**, 369–381 (1984)
42. Tadmor, E.: The numerical viscosity of entropy stable schemes for systems of conservation laws, i. *Math. Comp.* **49**, 91–103 (1987)
43. Tadmor, E.: Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numer.* pp. 451–512 (2003)
44. Toro, E.F.: *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer Verlag, Berlin (1999)
45. Toro, E.F.: *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Springer, Berlin (2009)
46. Trefethen, L.N., Trefethen, A.E., Reddy, S.C., Driscoll, T.A.: Hydrodynamic stability without eigenvalues. *Science* **261**(5121), 578–584 (1993). <https://doi.org/10.1126/science.261.5121.578>
47. Winters, A.R., Derigs, D., Gassner, G.J., Walch, S.: A uniquely defined entropy stable matrix dissipation operator for high Mach number ideal MHD and compressible Euler simulations. *J. Comput. Phys.* **332**, 274–289 (2017)
48. Yamaleev, N., Carpenter, M.: Third-order energy stable WENO scheme. *JCP* **228**(8), 3025–3047 (2009)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.