SIAM J. NUMER. ANAL.
Vol. 53, No. 4, pp. 1857–1875

ACCURACY OF HIGH ORDER AND SPECTRAL METHODS FOR HYPERBOLIC CONSERVATION LAWS WITH DISCONTINUOUS SOLUTIONS*

J. ZUDROP† AND J. S. HESTHAVEN‡

Abstract. Higher order and spectral methods have been used with success for elliptic and parabolic initial and boundary value problems with smooth solutions. On the other hand, higher order methods have been applied to hyperbolic problems with less success, as higher order approximations of discontinuous solutions suffer from the Gibbs phenomenon. We extend past work and show that spectral methods yield spectral convergence of moments, even when applied to problems with discontinuous solutions. Besides spectral Fourier methods for periodic domains we also prove high order convergence for adjoint-consistent numerical methods, exemplified by the discontinuous Galerkin finite element method.

Key words. spectral methods, discontinuous Galerkin, hyperbolic conservation law, Gibbs oscillations

AMS subject classifications. 65M70, 65M60, 76M22

DOI. 10.1137/140992758

1. Introduction.

High order and spectral methods have been used extensively for solving elliptic and parabolic problems with smooth solutions. Among the most well-known methods is the (pseudo-)spectral Fourier–Galerkin method [15, 12, 2, 4] on periodic domains, as a numerical tool for the study of turbulence in the incompressible Navier–Stokes equation. The most appealing property of such methods is their pointwise exponential error convergence, whenever the solution is smooth. In recent decades higher order and spectral methods have been generalized to structured and fully unstructured meshes [3, 20] with the goal to recover high order error convergence rates for problems in complex geometries.

In contrast, hyperbolic conservation laws with discontinuous solutions have been targeted less by high order and spectral methods as the approximation of a nonsmooth solution suffers from the Gibbs phenomenon: Pointwise convergence of an Nth order scheme is degraded to $\mathcal{O}(1/N)$ globally. Inside the $1/N$-interval around the point of discontinuity the solution is disturbed by $\mathcal{O}(1)$ oscillations.

However, Lax, Gottlieb, Maday, Tadmor, Shu, and others [22, 14, 23, 25] have argued that even though pointwise convergence is lost, higher order information is still inherently available. By using postprocessing techniques [19, 28, 27, 13] it may thus be possible to extract higher order accurate information from a solution affected by the Gibbs phenomenon. Furthermore, with the postprocessing techniques discussed in [18, 17] it is possible to reconstruct a pointwise accurate solution (with exponentially

*Received by the editors October 27, 2014; accepted for publication (in revised form) May 4, 2015; published electronically August 4, 2015.
http://www.siam.org/journals/sinum/53-4/99275.html

†Applied Supercomputing in Engineering, German Research School for Simulation Sciences, RWTH Aachen, Aachen 52062, Germany, and Simulation Techniques and Scientific Computing, University of Siegen, Siegen 57076, Germany (j.zudrop@grs-sim.de, jens.zudrop@uni-siegen.de). The research of this author was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) in the framework of the HPC software initiative in the project HISEEM.

‡MCSS, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland (jan.hesthaven@epfl.ch).

decaying error) up to the point of discontinuity, once the location of the discontinuity is known. If accuracy is indeed maintained, postprocessing of spectral and high order methods seems to remedy all deficiencies. Unfortunately little is known about the way spectral methods recover spectrally accurate information. For more complex problems it is unclear whether the postprocessing techniques [19, 28, 27, 13, 18, 17] suffice to recover this information.

Only a few rigorous results on the spectrally accurate information for general problems are available: for a linear hyperbolic PDE with smooth coefficients and a discontinuous initial condition, Abarbanel, Gottlieb, and Tadmor [1] showed that any moment can be recovered with spectral accuracy. However, little is known for nonlinear PDEs, even though numerical experiments [23, 11, 25] indicate that postprocessing techniques improve the solution quality and higher order pointwise convergence can be recovered.

In this work, we are concerned with the rigorous analysis of spectral methods for linear hyperbolic conservation laws where the discontinuity of the solution is imposed by a discontinuous advection coefficient, a nonsmooth reaction term, or a discontinuous source term. We prove that while moments of the exact solution are not spectrally accurate, highly accurate information is available and can be recovered with spectral accuracy, provided the information is properly extracted. We show that this result generalizes to stable, adjoint-consistent discretizations. In particular, we investigate high order discontinuous Galerkin finite element discretizations.

The outline is as follows. In section 2 we consider the spectral Fourier–Galerkin method. Section 2.1 briefly reviews the classical result of [1] for discontinuous initial conditions. In sections 2.2 and 2.3 we prove that higher order information is not destroyed by the spectral Fourier method for a linear conservation law with a discontinuous source term or discontinuous advection-reaction coefficient, respectively. A generalization of the results to other, nonperiodic, spectral methods is given in section 3, where we consider the discontinuous Galerkin finite element method (DGFEM). We show in section 3.1 that an adjoint-consistent discontinuous Galerkin discretization allows us to recover a similar convergence result. In section 4 we use nonlinear stability results of the DGFEM to generalize our main results to nonlinear scalar conservation laws. Section 5 presents several numerical experiments, and we give a conclusion and a short outlook to future work in section 6.

2. Fourier method. We consider the linear conservation law

$$\partial_t u(x,t) + \partial_x \left(L(x)u(x,t)\right) = 0 \tag{2.1}$$

with the hyperbolic differential operator $\mathcal{L}(x) = \partial_x L(x)$ on the space-time domain $\Omega \times (0;T]$ with initial conditions

$$u(x,0) = u_0(x). \tag{2.2}$$

Without loss of generality we consider a one-dimensional periodic domain Ω in $\mathbf{R}$. The extension to a multidimensional periodic $\Omega \subset \mathbf{R}^d$ is straightforward by a d-dimensional tensor product. The semidiscrete Fourier–Galerkin approximation of (2.1) is given by

$$\left\langle \partial_t u_N(x,t) + \partial_x \left(L(x)u_N(x,t)\right), \exp\left(ikx\right)\right\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N, \tag{2.3}$$

where $\langle\cdot\rangle_{L_2(\Omega)}$ denotes the L_2-scalar product on Ω and the Nth order solution is represented as

$$u_N(x,t) = \sum_{|p|\le N} \hat{u}_p(t) \exp\left(ipx\right). \tag{2.4}$$

The initial condition $u_N(0)$ is obtained by a simple L_2-projection,

$$\langle u_N(x,0) - u_0(x), \exp(ikx)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N. \tag{2.5}$$

Remark 2.1. A slightly different Fourier–Galerkin method for the linear conservation law (2.1) can be written as

$$\langle \partial_t u_N(x,t) + \partial_x \left(L_N(x) u_N(x,t)\right), \exp(ikx)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N, \tag{2.6}$$

where we have a discrete operator L_N obtained by an L_2-projection,

$$\langle L_N(x) - L(x), \exp(ikx)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N. \tag{2.7}$$

Even though it is a nonstandard Galerkin formulation, the results of this work hold true for this formulation, too.

Before we discuss the technical details, let us briefly outline the four key points of our results:

(i) In addition to (2.1), we define an auxiliary hyperbolic equation. This auxiliary equation is closely related to the adjoint problem.

(ii) The original equation and its auxiliary problem, as well as their semidiscrete versions, satisfy Green's identity, i.e., moments between the two solutions, either on the continuous or on the semidiscrete level, are conserved in time.

(iii) We show that the numerical approximations of the initial moments are spectrally accurate. By applying Green's identity we conclude that this information is conserved as we evolve the original equation and its auxiliary problem forward in time. This holds true in the continuous and semidiscrete setting.

(iv) To prove that information about any arbitrary $f \in C^\infty$ moment is available with spectral accuracy, we take advantage of the hyperbolic nature of the auxiliary problem. This allows us to solve the continuous auxiliary problem reverse in time and to solve the semidiscrete auxiliary problem forward in time to recover the spectrally accurate reconstructed solution.

2.1. Smooth coefficients and discontinuous initial condition. Here we briefly review the results of [1] for discontinuous initial condition $u_0(x) \in L_2(\Omega)$ and smooth coefficient $L(x) \in C^\infty(\Omega)$. This serves as a starting point of the following discussions.

We define the auxiliary problem

$$\partial_t v(x,t) + L(x)\partial_x v(x,t) = 0 \tag{2.8}$$

with initial condition

$$v(x,0) = v_0(x). \tag{2.9}$$

Notice that the auxiliary problem (2.8) is related to the adjoint operator $\mathcal{L}^* = -L(x)\partial_x$ of (2.1) by

$$\partial_t v(x,t) - \mathcal{L}^* v(x,t) = 0. \tag{2.10}$$

The semidiscrete Fourier–Galerkin formulation of the auxiliary problem (2.8) is given by

$$\langle \partial_t v_N(x,t) + L(x)\partial_x v_N(x,t), \exp(ikx)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N, \tag{2.11}$$

where

$$\langle v_N(x,0) - v_0(x), \exp(ikx)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N, \tag{2.12}$$

where $v_N(x,t)$ is expressed in terms of a Fourier series with coefficients $\hat{v}_p(t)$ for all $|p| \le N$. As u_0 is a discontinuous function, the solution u of (2.1) is defined in a weak sense. The following relation for the conservation law (2.1) and the auxiliary problem (2.8) holds true.

LEMMA 2.1 (continuous Green's identity). *The solutions u, v of* (2.1), (2.8) *satisfy*

$$\langle u(x,t), v(x,t)\rangle_{L_2(\Omega)} = \langle u_0(x), v_0(x)\rangle_{L_2(\Omega)} \quad \forall t \in (0;T]. \tag{2.13}$$

Proof. Using integration by parts we obtain for any $t \in (0;T]$

$$\begin{aligned}\partial_t \langle u(x,t), v(x,t)\rangle_{L_2(\Omega)} &= -\langle \partial_x \left(L(x)u(x,t)\right), v(x,t)\rangle_{L_2(\Omega)} - \langle u(x,t), L(x)\partial_x v(x,t)\rangle_{L_2(\Omega)} \\ &= -\langle \partial_x \left(L(x)u(x,t)\right), v(x,t)\rangle_{L_2(\Omega)} \\ &\quad + \langle \partial_x \left(L(x)u(x,t)\right), v(x,t)\rangle_{L_2(\Omega)} \\ &= 0. \qquad \square\end{aligned}$$

Similarly, the corresponding semidiscrete identity holds.

LEMMA 2.2 (semidiscrete Green's identity). *The solutions u_N, v_N of the semidiscrete Fourier–Galerkin methods* (2.3), (2.11) *satisfy*

$$\langle u_N(x,t), v_N(x,t)\rangle_{L_2(\Omega)} = \langle u_N(x,0), v_N(x,0)\rangle_{L_2(\Omega)} \quad \forall t \in (0;T]. \tag{2.14}$$

Proof. As u_N and v_N are trigonometric polynomials of order N and by construction of the Fourier–Galerkin method, they satisfy

$$\langle \partial_t u_N(x,t) + \partial_x \left(L(x)u_N(x,t)\right), v_N(x,t)\rangle_{L_2(\Omega)} = 0 \tag{2.15}$$

and

$$\langle u_N(x,t), \partial_t v_N(x,t) + L(x)\partial_x v_N(x,t)\rangle_{L_2(\Omega)} = 0. \tag{2.16}$$

Hence, after integration by parts we observe

$$\partial_t \langle u_N(x,t), v_N(x,t)\rangle_{L_2(\Omega)} = 0 \tag{2.17}$$

and the lemma follows. $\square$

The approximation error of the initial moment can be bounded as follows.

LEMMA 2.3. *Let $u_0(x) \in L_2(\Omega)$ and $v_0(x) \in C^\infty(\Omega)$ and let $u_N(x,0)$ and $v_N(x,0)$ be the Nth order L_2-approximant (given by* (2.5) *and* (2.12)*). Then the estimate*

$$|\langle u_N(x,0), v_N(x,0)\rangle - \langle u_0(x), v_0(x)\rangle| \le C\frac{\|v_0^{(s)}\|_{L_2(\Omega)}}{N^s} \sim C\|v_0\|_{L_2(\Omega)}\frac{s!}{N^s} \tag{2.18}$$

holds for $s \in \mathbf{N}$; the constant C depends on $u_0(x)$ but is independent of N. Hence, initial moments are exponentially accurate for $N \to \infty$.

Proof. As $v_N(0)$ is an Nth order polynomial we have

$$|\langle u_N(x,0) - u_0(x), v_N(x,0)\rangle_{L_2(\Omega)}| = 0 \tag{2.19}$$

by Galerkin orthogonality. By using Cauchy–Schwarz we have

$$\begin{aligned} |\langle u_0(x), v_N(x,0) - v_0(x)\rangle_{L_2(\Omega)}| &\le C\|u_0\|_{L_2(\Omega)}\|v_N - v_0\|_{L_2(\Omega)} \\ &\le C\|u_0\|_{L_2(\Omega)}\|v_0^{(s)}\|_{L_2(\Omega)}/N^s. \end{aligned} \tag{2.20}$$

Combining the two previous estimates we can conclude

$$\begin{aligned} |\langle u_N(x,0), v_N(x,0)\rangle - \langle u_0(x), v_0(x)\rangle| &= |\langle u_N(x,0) - u_0(x), v_N(x,0)\rangle \\ &\quad + \langle u_0(x), v_N(x,0) - v_0(x)\rangle| \\ &\le C\|u_0\|_{L_2(\Omega)}\|v_0^{(s)}\|_{L_2(\Omega)}/N^s \\ &\le C\|u_0\|_{L_2(\Omega)}\|v_0\|_{L_2(\Omega)}s!/N^s, \end{aligned} \tag{2.21}$$

which completes the proof. ∎

The three previous lemmas can be combined to derive the main theorem [1].

THEOREM 2.4. *Let $u(x,t), v(x,t)$ be the exact solution of* (2.1) *and* (2.8)*. Furthermore, let $u_N(x,t), v_N(x,t)$ be the semidiscrete solutions of the Fourier–Galerkin methods* (2.3), (2.11) *and let $v_0(x)$ be an arbitrary function in $C^\infty(\Omega)$. Then the following estimate holds for $t \in (0;T]$ and $s \in \mathbf{N}$:*

$$|\langle u_N(x,t), v(x,t)\rangle_{L_2(\Omega)} - \langle u(x,t), v(x,t)\rangle_{L_2(\Omega)}| \sim C_2\|v_0\|_{L_2(\Omega)}\frac{s!}{N^s}, \tag{2.22}$$

i.e., the moments of u_N (calculated with respect to v) are exponentially accurate.

Proof. Since $v_0(x) \in C^\infty(\Omega)$ and the adjoint problem (2.8) has smooth coefficients, we have that $v(x,t) \in C^\infty(\Omega)$ for $t \in (0;T]$. Hence, we can replace v by v_N in the inner products as $N \to \infty$. Therefore, we obtain

$$\begin{aligned} |\langle u_N(x,t), v_N(x,t)\rangle_{L_2(\Omega)} - \langle u(x,t), v(x,t)\rangle_{L_2(\Omega)}| &= |\langle u_N(x,0), v_N(x,0)\rangle_{L_2(\Omega)} \\ &\quad - \langle u_0(x), v_0(x)\rangle_{L_2(\Omega)}| \sim C\|v_0\|_{L_2(\Omega)}s!/N^s, \end{aligned} \tag{2.23}$$

which completes the proof. ∎

Remark 2.2. The previous analysis also applies for initial conditions with finite singularities, such that the methods presented in [5] can be applied to recover pointwise accurate results.

2.2. Discontinuous source term. We are now concerned with a linear hyperbolic conservation law with smooth coefficients $L(x)$, zero initial condition u_0, and a discontinuous source term $f \in L_2(\Omega)$,

$$\partial_t u(x,t) + \partial_x\left(L(x)u(x,t)\right) = f(x,t) \tag{2.24}$$

in the space-time domain $\Omega \times (0;T]$.

The exact solution of (2.24) can be found by applying Duhamel's principle, i.e., by representing the solution of the inhomogeneous (2.24) by the solution of a homogeneous initial value problem. The solution $u(x,t)$ is given by

$$u(x,t) = \int_0^t P^\alpha(f)(x,t)d\alpha, \tag{2.25}$$

where $P^\alpha(f)(x,t) = w^\alpha(x,t)$ is the solution operator to the problem

$$\partial_t w^\alpha + \partial_x\left(L(x)w^\alpha(x,t)\right) = 0 \tag{2.26}$$

with initial condition

$$w^\alpha(x,\alpha) = f(x,\alpha) \tag{2.27}$$

on the space-time domain $\Omega \times (\alpha; T]$. Notice that the solutions u, w to (2.24) and (2.26) are defined in a weak sense.

The Nth order Fourier–Galerkin approximation to (2.24) reads

$$\langle \partial_t u_N(x,t) + \partial_x \left(L(x)u_N(x,t)\right) - f(x,t), \exp\left(ikx\right)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N, \tag{2.28}$$

and the discrete initial condition is obtained by (2.5).

LEMMA 2.5 (discrete Duhamel's priniciple). *The semidiscrete solution $u_N(x,t)$ of the Fourier–Galerkin method is given by*

$$u_N(x,t) = \int_0^t P_N^\alpha(f)(x,t)d\alpha, \tag{2.29}$$

where $P_N^\alpha(f)(x,t) = w_N^\alpha(x,t)$ is the solution operator to the semidiscrete homogeneous initial value problem

$$\langle \partial_t w_N^\alpha(x,t) + \partial_x \left(L(x)w_N^\alpha(x,t)\right), \exp\left(ikx\right)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N, \tag{2.30}$$

$$\langle w_N^\alpha(x,\alpha) - f(x,\alpha), \exp\left(ikx\right)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N \tag{2.31}$$

on $\Omega \times (\alpha, T]$.

Proof. The proof follows by direct calculation. ∎

This discrete version of Duhamel's principle allows us to use the results of section 2.1, applied to the solutions w^α, w_N^α of the homogeneous initial values problems for any $\alpha \in (0; T]$.

THEOREM 2.6. *Let u be the exact solution of (2.24) and u_N be the semidiscrete solution obtained by the Fourier–Galerkin method (2.28). Furthermore, let v be an arbitrary function in $C^\infty(\Omega)$. Then the error estimate*

$$|\langle u_N(x,t), v(x,t)\rangle - \langle u(x,t), v(x,t)\rangle| \le C_3 \frac{\|v^{(s)}\|_{L_2(\Omega)}}{N^s} \sim C_3 \|v\|_{L_2(\Omega)} \frac{s!}{N^s} \tag{2.32}$$

holds for $s \in \mathbf{N}$ and $N \to \infty$. The constant C_3 is independent of N. Hence we recover exponential accuracy for moments of u.

Proof. For any $\alpha \in (0; T]$, we define an auxiliary equation by

$$\partial_t v^\alpha(x,t) + L(x,t)\partial_x v^\alpha(x,t) = 0 \tag{2.33}$$

and for the semidiscrete problem as

$$\langle \exp\left(ikx\right), \partial_t v_N^\alpha(x,t) + L(x,t)\partial_x v_N^\alpha(x,t)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N. \tag{2.34}$$

By solving the hyperbolic equation backward in time, such that $v^\alpha(x,t) = v \in C^\infty(\Omega)$, we obtain an initial condition $v(x,\alpha) = v_0(\alpha, x) \in C^\infty(\Omega)$ (smoothness of v_0 follows from $L(x) \in C^\infty(\Omega)$). For each $\alpha \in (0; T]$ we apply Theorem 2.4 and deduce

$$|\langle P_N^\alpha(f_N)(x,t), v(x,t)\rangle - \langle P^\alpha(f)(x,t), v(x,t)\rangle| \le C \frac{\|v_0^{(s)}\|_{L_2(\Omega)}}{N^s} \sim C\|v_0\|_{L_2(\Omega)} \frac{s!}{N^s}. \tag{2.35}$$

The constant C is independent of N and bounded from above in terms of α. The theorem follows by applying the previous results for (2.25) and (2.29). ∎

2.3. Conservation law with discontinuous advection-reaction coefficients. In this section we consider a linear hyperbolic conservation law with a reaction term

$$\partial_t u(x,t) + \partial_x \left(L(x)u(x,t)\right) + \sigma(x)u(x,t) = 0 \tag{2.36}$$

in the space-time domain $\Omega \times (0;T]$. We assume that $L(x), \sigma(x) \in L_2(\Omega)$ are discontinuous coefficient functions. Furthermore, we assume that the initial condition is smooth,

$$u(x,0) = u_0(x) \in C^\infty(\Omega). \tag{2.37}$$

Let us define an auxiliary equation by

$$\partial_t v(x,t) + L(x)\partial_x v(x,t) - \sigma(x)v(x,t) = 0 \tag{2.38}$$

with initial condition

$$v(x,0) = v_0(x) \in L_2(\Omega). \tag{2.39}$$

Notice that the differential operator $L(x)\partial_x$ is hyperbolic, provided the original differential operator $\partial_x L(x)$ in (2.36) is hyperbolic. As the coefficients $L(x), \sigma(x)$ are discontinuous functions, we define the solutions to (2.36) and (2.38) in a weak sense.

LEMMA 2.7. *Let u, v be the exact solution of* (2.36) *and* (2.38)*, respectively. Then, for any $t \in (0;T]$*

$$\langle u(x,t), v(x,t)\rangle_{L_2(\Omega)} = \langle u_0(x), v_0(x)\rangle_{L_2(\Omega)} \tag{2.40}$$

holds true.

Proof. The lemma follows by using integration by parts; cf. Lemma 2.1. ∎

Similar to the previous sections, we define the Nth order Fourier–Galerkin scheme for (2.36) and (2.38) as

$$\langle \partial_t u_N(x,t) + \partial_x \left(L(x)u_N(x,t)\right) + \sigma(x)u_N(x,t), \exp\left(ikx\right)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N \tag{2.41}$$

and

$$\langle \exp\left(ikx\right), \partial_t v_N(x,t) + L(x)\partial_x v_N(x,t) - \sigma(x)v_N(x,t)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N. \tag{2.42}$$

LEMMA 2.8. *Let u_N, v_N be the numerical solutions of the Nth order Fourier (–Galerkin) methods* (2.41) *and* (2.42)*. Then for any $t \in (0;T]$*

$$\langle u_N(x,t), v_N(x,t)\rangle_{L_2(\Omega)} = \langle u_N(x,0), v_N(x,0)\rangle_{L_2(\Omega)} \tag{2.43}$$

holds true.

Proof. The proof of this lemma is similar to the proof of Lemma 2.2. ∎

The two previous lemmas give rise to the main result of this section. We show that high order information is still contained in the numerical data and it can be recovered.

THEOREM 2.9. *Let $u_0 \in C^\infty(\Omega)$ and u_N be the Nth order Fourier–Galerkin solution of* (2.41)*, and let u be the exact solution of* (2.36)*. Furthermore, let $f \in C^\infty(\Omega)$ be an arbitrary smooth function. As $N \to \infty$, there exists a sequence of functions $\{g_N\}$ (where $g_N \in \mathbf{P}_N(\Omega)$ is a trigonometric polynomial of order N) with*

$$\lim_{N\to\infty} \|g_N - f\|_{L_2(\Omega)} \to 0 \tag{2.44}$$

and for any $s \in \mathbf{N}$
(2.45)

$$|\langle u_N(x,T), g_N(x)\rangle_{L_2(\Omega)} - \langle u(x,T), f(x)\rangle_{L_2(\Omega)}| \leq C\frac{\|u_0^{(s)}\|_{L_2(\Omega)}}{N^s} \sim C\|u_0\|_{L_2(\Omega)}\frac{s!}{N^s},$$

i.e., exponentially accurate information about moments is contained in the numerical data. Furthermore, the functions g_N *are uniquely determined, i.e.,* $g_N = \Gamma_N(f)$*, with some mapping* $\Gamma_N : C^\infty(\Omega) \to \mathbf{P}_N(\Omega)$*.*

Proof. Let $f(x) = v(x,T)$ be an arbitrary function in $C^\infty(\Omega)$. Solve the auxiliary problem (2.38) reverse in time and find $v_0(x)$. Notice that this could be done due to the hyperbolic nature of the differential operator $L(x)\partial_x$. However, due to the low regularity of the coefficient function $L(x)$, the corresponding solution v_0 of the backward problem at $t = 0$ will also be of low regularity.

Now, we use v_0 as the initial condition for the Fourier–Galerkin method (2.42) of the adjoint problem, obtained by

$$\langle \exp(ikx), v_N(x,0) - v_0(x)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \leq N, \tag{2.46}$$

and solve the problem forward in time to T. We set $g_N(x) = v_N(x,T)$. As L, σ are discontinuous functions, its numerical approximations L_N, σ_N are not spectrally accurate. Nevertheless, they converge with first order accuracy in L_2, and hence we have $\|g_N - f\|_{L_2(\Omega)} \sim 1/N$. Obviously, the procedure above defines a well-posed mapping $\Gamma_N : C^\infty(\Omega) \to \mathbf{P}_N(\Omega)$.

Since $u_N(x,0)$ is a trigonometric polynomial of order N, we have

$$\langle u_N(x,0), v_N(x,0) - v_0(x)\rangle_{L_2(\Omega)} = 0 \tag{2.47}$$

by Galerkin orthogonality. Since $u_0 \in C^\infty(\Omega)$, its L_2-projection $u_N(0)$ is a spectrally accurate approximation. Combined with (2.47) we have that for $N \to \infty$ and $s \in \mathbf{N}$

$$|\langle u_N(x,0), g_N(x,0)\rangle_{L_2(\Omega)} - \langle u_0(x), v_0(x)\rangle_{L_2(\Omega)}| \leq C\frac{\|u^{(s)}\|_{L_2(\Omega)}}{N^s}, \tag{2.48}$$

where C is independent of N. Hence the approximation of moments of the initial conditions is spectrally accurate, even though the initial condition v_0 of the adjoint problem is a discontinuous function. The theorem follows by applying Lemmas 2.7 and 2.8 to (2.48). ∎

We conclude this section with two short remarks regarding the previous theoretical considerations.

Remark 2.3. In contrast to the results of section 2.1, where the initial condition was discontinuous, the term

$$|\langle u_N(x,T), f(x)\rangle_{L_2(\Omega)} - \langle u(x,T), f(x)\rangle_{L_2(\Omega)}| \tag{2.49}$$

is in general not exponentially small as $N \to \infty$, i.e., it is of order $\mathcal{O}(1/N)$. Nevertheless, along the line $\theta f + (1-\theta)g_N$ with $\theta \in [0;1]$, accuracy of the moments improves continuously as $\theta \to 0$. In this sense g_N and f are connected.

Remark 2.4. Notice that the results above apply not only to scalar, linear hyperbolic conservation laws but also to systems of conservation laws in multiple spatial dimensions. Hence, it shows that even counterpropagating waves do not destroy high order information of the moments in a numerical solution obtained by the Fourier–Galerkin method.

In the following section we consider nonperiodic numerical methods and show how the results of section 2 can be generalized to more generic setups. In particular, we consider the DGFEM and prove that the previous results apply here as well.

3. Discontinuous Galerkin method. The DGFEM has emerged as a popular numerical method for hyperbolic conservation laws over the last decade. Originally, it was introduced by Reed and Hill in [24] to solve the steady-state neutron transport equation and subsequently it was analyzed and developed extensively, in particular through a series of papers by Cockburn and Shu [8, 7, 6, 9]. Today, it is applied to complex linear and nonlinear problems, as well as to high order equations and equations in nonconservative forms. For a full review of such developments, we refer the reader to [20, 10].

Let us briefly recall the basic formulation of the DGFEM for a linear system of hyperbolic conservation laws with source terms, as given in (2.36). We start from a tesselation $\Upsilon = \{\Omega_i | i = 1, \ldots, n\}$ of the domain Ω into a set of n nonoverlapping elements Ω_i. On the tesselation we define the broken Sobolev space $W^{k,p}(\Upsilon) = \{f : \Omega \to \mathbf{R} \in L^p(\Omega) \forall \Omega_i \in \Upsilon : f|_{\Omega_i} \in W^{k,p}(\Omega_i)\}$ and the broken polynomial space of order N order $\mathbf{P}^N(\Upsilon) = \{f : \Omega \to \mathbf{R} \forall \Omega_i \in \Upsilon : f \in \mathbf{P}^N(\Omega_i)\}$. The variational formulation is obtained by multiplying (2.36) with a test function, chosen from a properly defined space, denoted by $X(\Omega)$ in the following, integrating by parts in each element and replacing the surface integrals by properly defined numerical flux functions. Overall, we arrive at the following variational formulation of (2.36): Find $u \in L_2(\Omega)$ such that for all $\psi \in X(\Omega)$

$$\sum_{i=1}^{n} \partial_t \int_{\Omega_i} u\psi dV - \int_{\Omega_i} Lu\partial_x \psi dV + \int_{\partial\Omega_i} L^*(u^+, u^-, n)\psi dS + \int_{\Omega_i} \sigma u\psi dV = 0. \tag{3.1}$$

The numerical flux $L^*(u^+, u^-, n)$ is a numerical approximation to the flux Lun in Ω_i's outward outer unit normal direction n. It is a function of the inner trace u^- of Ω_i on $\partial\Omega_i$ and the outer trace u^+.

Finally, the semidiscrete DGFEM formulation is obtained by restricting the ansatz and test functions space to the finite-dimensional space $\mathbf{P}^N(\Upsilon)$. Find $u_N \in \mathbf{P}^N(\Upsilon)$ such that for all $\psi_N \in \mathbf{P}^N(\Upsilon)$

$$\sum_{i=1}^{n} \partial_t \int_{\Omega_i} u_N\psi_N dV - \int_{\Omega_i} Lu_N\partial_x \psi_N dV + \int_{\partial\Omega_i} L^*(u_N^+, u_N^-, n)\psi_N dS + \int_{\Omega_i} \sigma u_N\psi_N dV = 0. \tag{3.2}$$

For convenience, we write the semidiscrete scheme (3.2) in terms of the bilinear forms $\mathcal{M}, \mathcal{B}$, i.e., find $u_N \in \mathbf{P}^N(\Upsilon)$ such that

$$\partial_t \mathcal{M}(u_N, \psi_N) + \mathcal{B}(u_N, \psi_N) = 0 \quad \forall \psi_N \in \mathbf{P}^N(\Upsilon). \tag{3.3}$$

3.1. Adjoint-consistency. In this section, we prove that adjoint-consistency and stability are sufficient conditions to ensure that the results of section 2 carry over to the discontinuous Galerkin method. We start by defining adjoint-consistency for smooth hyperbolic problems.

DEFINITION 3.1 (adjoint-consistency). *Let L, σ be in $\mathbf{P}^N(\Upsilon)$. The discontinuous Galerkin approximation* (3.3) *of* (2.36) *is adjoint-consistent if*

$$\partial_t \mathcal{M}(\phi_N, v) + \mathcal{B}(\phi_N, v) = 0 \quad \forall \phi_N \in \mathbf{P}^N(\Upsilon) \tag{3.4}$$

is satisfied for all $t \in (0; T]$ by the exact solution $v(x, t)$ of the adjoint problem

$$\partial_t v(x, t) - L(x)\partial_x v(x, t) + \sigma(x)v(x, t) = 0. \tag{3.5}$$

The previous definition provides a variational formulation of the numerical scheme for the auxiliary problem (2.38), i.e., find $v_N \in \mathbf{P}^N(\Upsilon)$ such that

$$\partial_t \mathcal{M}(\phi_N, v_N) - \mathcal{B}(\phi_N, v_N) = 0 \quad \forall \phi_N \in \mathbf{P}^N(\Upsilon). \tag{3.6}$$

A number of discontinuous Galerkin formulations are adjoint-consistent. In the following we focus on central based formulations for simplicity.

LEMMA 3.2. *The DGFEM formulation* (3.3) *based on the central flux is adjoint-consistent.*

Proof. First we consider a simple scalar conservation law with central flux

$$L^*(u_N^+, u_N^-, n) = L\frac{u_N^+ + u_N^-}{2} \cdot n \tag{3.7}$$

and prove that this formulation is adjoint-consistent. A generalization to systems of conservation laws is straightforward. Inserting the flux into the DGFEM formulation we obtain, after regrouping the trace terms on the faces for all Ω_i, that

$$\sum_{i=1}^{n} \int_{\Omega_i} u_N \big(\partial_t \psi_N - L\partial_x \psi_N + \sigma \psi_N\big) dV + \int_{\partial\Omega_i} u_N L\frac{\psi_N^+ - \psi_N^-}{2} \cdot n dS = 0. \tag{3.8}$$

For $\psi_N = v$ the volume integral vanishes by definition. Furthermore, by checking the Rankine–Hugoniot condition for the solution v of (3.5), it is easy to check that the solution is continuous along its characteristics. In particular, it implies that

$$\int_{\partial\Omega_i} L(v^+ - v^-) \cdot n dS = 0 \tag{3.9}$$

for $i = 1, \ldots, n$. Notice that these arguments still hold true for systems of conservation laws. This shows

$$\partial_t \mathcal{M}(\phi_N, v) + \mathcal{B}(\phi_N, v) = 0 \quad \forall \phi_N \in \mathbf{P}^N(\Upsilon) \tag{3.10}$$

and the proof is complete. ∎

Now, we are able to state the main result of this section for adjoint-consistent discontinuous Galerkin methods.

THEOREM 3.3. *Theorems* 2.4 and 2.9 *carry over to stable, adjoint-consistent DGFEM if* (3.6) *is a consistent and stable discretization of* (2.8).

Proof. The proof for $L, \sigma \in C^\infty(\Omega)$ and discontinuous $u_0 \in L_2(\Omega)$ follows by the proof of Theorem 2.4, applying the auxiliary numerical method (3.6) and by applying basic error estimates of polynomial approximations in broken polynomial spaces [10].

For discontinuous $L, \sigma \in L_2(\Omega)$ and smooth u_0, we notice that L_N and σ_N are first order approximations as $N \to \infty$. Therefore, the numerical schemes (3.3) and (3.6) define first order schemes in terms of N. The remaining steps follow the proof of Theorem 2.9. ∎

Notice that (3.6) (implicitly defined by (3.8)) is a stable and consistent discretization of (2.8) and thus Theorem 3.3 applies for the discontinuous Galerkin discretization equipped with the flux (3.7). If pure upwinding is used in the discontinuous Galerkin formulation (3.3), the discrete auxiliary problem (3.6) is not stable (downwinding). Instead, one may directly apply upwinding to solve the discrete auxiliary

problem (3.6) forward in time. In this case, stability is guaranteed and one obtains, by utilizing the procedure from the proof of Theorem 3.3, the bound

$$|\langle u_N(x,T), g_N(x)\rangle_{L_2(\Omega)} - \langle u(x,T), f(x)\rangle_{L_2(\Omega)}| \le C\frac{\|u_0^{(s)}\|_{L_2(\Omega)}}{N^s} + C', \tag{3.11}$$

where

$$C' = \sum_i \int_0^T \int_{\partial\Omega_i, n\cdot L\ge 0} L(v_N^+ - v_N^-)u_N \cdot n dS + \int_{\partial\Omega_i, n\cdot L<0} (-L)(v_N^+ - v_N^-)u_N \cdot n dS dt. \tag{3.12}$$

Thus, if one can guarantee that C' is small, e.g., for smooth v with local support separated from the point of discontinuity, $\langle u_N(x,T), g_N(x)\rangle$ will converge rapidly toward $\langle u(x,T), f(x)\rangle$.

We finish this section with the following remark regarding simple mesh refinement, i.e., h-convergence (where $h \sim 1/n$), of an Nth order scheme (for fixed N).

Remark 3.1. The proof of Theorem 3.3 shows that for h-convergence of an Nth order scheme, high order accuracy is maintained, i.e., let u_N^h be the numerical solution on $n \sim 1/h$ elements of degree N, and let $f \in C^\infty(\Omega)$. If the numerical schemes (3.3), (3.6) are optimal order convergent, we obtain, for the discontinuous initial condition setup (with smooth coefficients), that

$$|\langle u_N^h, f\rangle_{L_2(\Omega)} - \langle u, f\rangle_{L_2(\Omega)}| \le Ch^{N+1} \tag{3.13}$$

for $h \to 0$. For the discontinuous coefficient problem with smooth initial condition we have

$$|\langle u_N^h, g_N^h\rangle_{L_2(\Omega)} - \langle u, f\rangle_{L_2(\Omega)}| \le Ch^{N+1}, \tag{3.14}$$

where g_N^h are the auxiliary numerical results on $n \sim 1/h$ elements of degree N, and $\|g_N^h - f\|_{L_2(\Omega)} \sim h$ for $h \to 0$.

4. Some remarks for nonlinear conservation laws. Weak solutions of nonlinear conservation laws

$$\partial_t u + \partial_x F(u) = 0 \tag{4.1}$$

require additional consideration as follows: Convergence of u_N toward the entropy solution u of the original equations is not always ensured. Let us give a small example to demonstrate the problem with smoothing procedures for spectral methods: if we consider the spectral-viscosity Fourier method [26] for Burgers equation $\partial_t u + \partial_x u^2/2 = 0$, we obtain

$$\langle \partial_t u_N + \partial_x\left(\frac{1}{2}u_N^2\right) - \partial_x \nu(I - P_m)\partial_x u_N, \exp(ikx)\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N \tag{4.2}$$

with $\nu = \nu(N)$ and $m = m(N)$ and P_m is the L_2-projection onto the first m Fourier modes. While it can be shown by compensated compactness arguments that u_N converges to the entropy solution, its formal auxiliary problem, given by

$$\langle \exp(ikx), \partial_t v_N + \frac{1}{2}u_N \partial_x v_N + \partial_x \nu(I - P_m)\partial_x v_N\rangle_{L_2(\Omega)} = 0 \quad \forall |k| \le N, \tag{4.3}$$

is antidissipative in nature and is therefore ill-posed.

The DGFEM is exceptional in the following sense: Jiang and Shu [21] showed that if the method converges (for arbitrary N), it converges to the entropy solution, provided a monotone numerical flux is used, and the flux is convex. Hence, the viscosity introduced by the numerical flux suffices to guarantee that the solution converges to the entropy solution. Since no additional dissipation is necessary, this allows us to maintain the hyperbolic framework introduced in the previous sections. In the following we focus on the Burgers equation for simplicity. However, the basic ideas are not limited to it. We assume for simplicity that u is either strictly positive or strictly negative.

The semidiscrete DGFEM formulation for Burgers equation is

(4.4)
$$\sum_{i=1}^{n} \partial_t \int_{\Omega_i} u_N \psi_N dV - \int_{\Omega_i} F(u_N, u_N) \partial_x \psi_N dV + \int_{\partial\Omega_i} F^*(u_N^+, u_N^+, u_N^-, u_N^-, n) \psi_N dS = 0.$$

For convenience, we write the flux F as the bilinear form $F(a,b) = ab/2$ and the Lax–Friedrich flux

$$F^*(a^+, b^+, a^-, b^-, n) = \frac{1}{2}\left(F(a^-, b^-) + F(a^+, b^+)\right) n + \alpha\left(a^- - a^+\right) \tag{4.5}$$

with α being an upper bound of the absolute wave-speed in the normal direction. For convenience we rewrite it in operator notation with the forms $\mathcal{M}, \mathcal{F}$,

$$\partial_t \mathcal{M}(u_N, \psi_N) + \mathcal{F}(u_N, u_N, \psi_N) = 0 \quad \forall \psi_N \in \mathbf{P}^N(\Upsilon). \tag{4.6}$$

The forms $\mathcal{M}, \mathcal{F}$ are defined as

$$\mathcal{M}(a,c) = \sum_{i=1}^{n} \int_{\Omega_i} acdV, \tag{4.7}$$

$$\mathcal{F}(a,b,c) = \sum_{i=1}^{n} -\int_{\Omega_i} F(a,b)\partial_x c dV + \int_{\partial\Omega_i} F^*(a^+, b^+, a^-, b^-, n) c dS. \tag{4.8}$$

Now, we consider adjoint-consistency for smooth solutions of the Burgers equation.

LEMMA 4.1. *The semidiscrete, Lax–Friedrich-based DGFEM formulation for the Burgers equation is adjoint-consistent, i.e., let u be the smooth solution of the Burgers equation, then*

$$\partial_t \mathcal{M}(\phi_N, v) + \mathcal{F}(\phi_N, u, v) = 0 \quad \forall \phi_N \in \mathbf{P}^N(\Upsilon) \tag{4.9}$$

for the exact (and smooth) solution $v(x,t)$ of

$$\partial_t v(x,t) - \frac{u(x,t)}{2} \partial_x v(x,t) = 0. \tag{4.10}$$

Proof. We select an arbitrary $\phi_N \in \mathbf{P}^N(\Upsilon)$ and set $a = \phi_N$, $b = u$, and $c = v$ in (4.7), (4.8). The volume integrals in (4.4), i.e., the sum of the volume integrals in (4.7), (4.8), vanish. To show that the numerical flux integral vanishes, we consider a unique face $\partial\kappa$ and consider the two flux terms from its left (superscript L) and right (superscript R) elements and sum them up:

(4.11)
$$\int_{\partial\kappa} \frac{1}{2}\left(F(u, \phi_N^L) + F(u, \phi_N^R)\right)\left(v^L - v^R\right) n^L dS + \int_{\partial\kappa} \alpha\left(\phi_N^L - \phi_N^R\right)\left(v^L - v^R\right) dS.$$

According to the Rankine–Hugoniot condition $v \in C^0(\Omega)$ along its characteristics (i.e., $\alpha(v^L - v^R) = 0$ and $un(v^L - v^R) = 0$). Hence, the proof is complete. ∎

Similar to section 3, the previous discussion gives rise to a numerical scheme for the auxiliary problem

$$\partial_t v(x,t) + \frac{u(x,t)}{2}\partial_x v(x,t) = 0 \tag{4.12}$$

to recover $v_N \in \mathbf{P}^N(\Upsilon)$ such that

$$\partial_t \mathcal{M}(\phi_N, v_N) - \mathcal{F}(\phi_N, u_N, v_N) = 0 \quad \forall \phi_N \in \mathbf{P}^N(\Upsilon). \tag{4.13}$$

THEOREM 4.2. *Let $u_0 \in C^\infty(\Omega)$ and u_N^h be the Nth order DGFEM solution on $1/h \sim n$ elements. Let u be the entropy solution and let $f \in C^\infty(\Omega)$. If u_N^h converges as $h \to 0$, it converges to the entropy solution and there exists a sequence of functions $\{g_N^h\}$ (where $g_N^h \in \mathbf{P}^N(\Upsilon)$) with*

$$\lim_{n\to\infty} \|g_N^h - f\|_{L_2(\Omega)} \to 0 \tag{4.14}$$

and

$$|\langle u_N(x,T), g_N^h(x)\rangle_{L_2(\Omega)} - \langle u(x,T), f(x)\rangle_{L_2(\Omega)}| \le Ch^{N+1} + C'. \tag{4.15}$$

The functions g_N^h are uniquely determined, i.e., $g_N^h = \Gamma_N^h(f, u_0)$, with some mapping $\Gamma_N^h : C^\infty(\Omega) \times C^\infty(\Omega) \to \mathbf{P}^N(\Upsilon)$. The constant C' is given by

$$C' = \sum_\kappa \int_0^T \int_{\partial\kappa} 2\alpha \left(u_N^L - u_N^R\right)\left(v_N^L - v_N^R\right) dSdt. \tag{4.16}$$

Hence, high order accurate information about moments is contained in the numerical data.

Proof. Due to the strictly hyperbolic nature of the conservation law, we observe that the auxiliary problem (4.12) is strictly hyperbolic, too. Hence, it can be solved exactly back in time. Then we apply a Lax–Friedrich-based approximation to the auxiliary problem (4.12) and solve it forth in time. The theorem follows by applying the local entropy results of [21] and by using arguments as in the proof of Theorem 3.3. Notice that $g_N^h = \Gamma_N^h(f, u_0)$ is now a function of f and u_0, as the auxiliary problem is dependent on the entropy solution u. ∎

We would like to remark that Theorem 4.2 does not automatically apply for nonlinear systems of conservation laws, as the entropy solution is, in general, not known to exist.

5. Numerical results. This section seeks to demonstrate the validity of the theoretical considerations through numerical experiments. The problems in sections 5.1–5.3 are solved by the Fourier–Galerkin method, while the problems in sections 5.4–5.5 are solved by the DGFEM. All semidiscrete systems are evolved forward in time by a classical fourth order Runge–Kutta method and we choose a sufficiently small time step to remove any temporal discretization error. Before we begin the discussion of the numerical results, we briefly recall the procedure of the numerical approach, which can be decomposed into the following steps:

(i) Solve the original problem forward in time and obtain $u(x,T)$.

(ii) Solve the auxiliary partial differential equation backward in time with given $f(x) = v(x,T) \in C^\infty(\Omega)$ and obtain $v(x,0)$.

(iii) Use $v_0(x) = v(x,0)$ to advance the discrete auxiliary problem forward in time and obtain $g_N(x) = v_N(x,T) \in \mathbf{P}^N$.

(iv) Compute $\langle u_N, f\rangle_{L_2(\Omega)}$, which we refer to as the smooth moment, compute $\langle u_N, g_N\rangle_{L_2(\Omega)}$, which we refer as the oscillatory moment, and consider convergence of the two quantities toward the exact value $\langle u, f\rangle_{L_2(\Omega)}$.

We would like to emphasize that this procedure is a proof of concept only. In realistic applications, where the reverse auxiliary problem cannot be solved exactly, postprocessing techniques like [19, 28, 27, 13, 18, 17] have to be applied. Furthermore, it should be noted that the previous procedure requires a stable discretization of the original and auxiliary problems. Stability of spectral Fourier–Galerkin methods for variable coefficient problems has been shown in [16]. For the discontinuous Galerkin method stability has been considered, for example, in [20].

5.1. Advection equation with discontinuous coefficient. We consider the one-dimensional scalar advection equation $\partial_t u + \partial_x a u = 0$ on the space-time domain $[0; 2\pi] \times (0; 1]$. The advection coefficient is given by

$$a(x) = \begin{cases} 1 & \text{if } x \le 2\pi/3, \\ 0.5 & \text{else,} \end{cases} \tag{5.1}$$

and the initial condition is set to

$$u_0(x) = \cos(5x + 0.1) + 0.1\cos(9x) + 0.001\sin(20x) + 0.0001\sin(44x). \tag{5.2}$$

Now, we investigate the convergence of moments for the numerical solutions u_N toward the the exact moments of u. In the following we set $f(x) = \sin(x)$; however, any other smooth function is possible. The resulting moment-error plots are shown in Figure 1. The left plot shows the (raw) error of the first 10 Fourier coefficients; the error of the Fourier coefficients decays with $1/N$ as $N \to \infty$. The right plot shows the convergence of the moment $\langle u, f\rangle_{L_2(\Omega)}$ at $T = 1$. The smooth moment $\langle u_N, f\rangle_{L_2(\Omega)}$ converges only like $1/N$ toward $\langle u, f\rangle_{L_2(\Omega)}$, while the oscillatory moment $\langle u_N, g_N\rangle_{L_2(\Omega)}$ converges exponentially fast toward $\langle u, f\rangle_{L_2(\Omega)}$ as $N \to \infty$. Overall, the plots are in perfect agreement with the theoretical predictions of section 2.

5.2. Reaction equation with discontinuous coefficient. Now consider the reaction equation $\partial_t u + \sigma u = 0$ on the space-time domain $[0; 2\pi] \times (0; 1]$ with the initial condition

$$u_0(x) = \cos(5x + 0.1) + 0.1\cos(9x) + 0.001\sin(20x) + 0.0001\sin(44x) \tag{5.3}$$

and the reaction coefficient

$$\sigma(x) = \begin{cases} 1 & \text{if } x \le 2\pi/3, \\ 0 & \text{else.} \end{cases} \tag{5.4}$$

Figure 2 shows the resulting error for the moments of u_N; the left plot shows the error of the first 10 Fourier coefficients, while the right plot illustrates the convergence toward the exact value $\langle u, f\rangle_{L_2(\Omega)}$ with $f = \sin(x)$. Again we observe slow $1/N$ convergence for the smooth moments, while the oscillatory moment is exponentially accurate as $N \to \infty$.

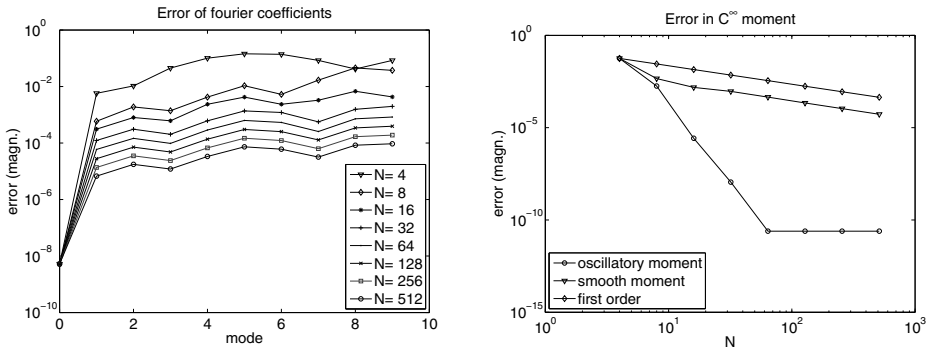

FIG. 1. *Error for $C^\infty(\Omega)$ moments of advection equation test case in section* 5.1 *at time* $t = 1.0$. *The left plot shows the error of the first 10 Fourier coefficients for various orders. The right plot shows the error for the $C^\infty(\Omega)$ moment $\langle u, \sin\rangle$. Smooth moment refers to $|\langle u_N, \sin\rangle - \langle u, \sin\rangle|$, while oscillatory moment refers to $|\langle u_N, g_N\rangle - \langle u, \sin\rangle|$. The smooth moment converges only as $1/N$, while the oscillatory moment converges exponentially fast.*

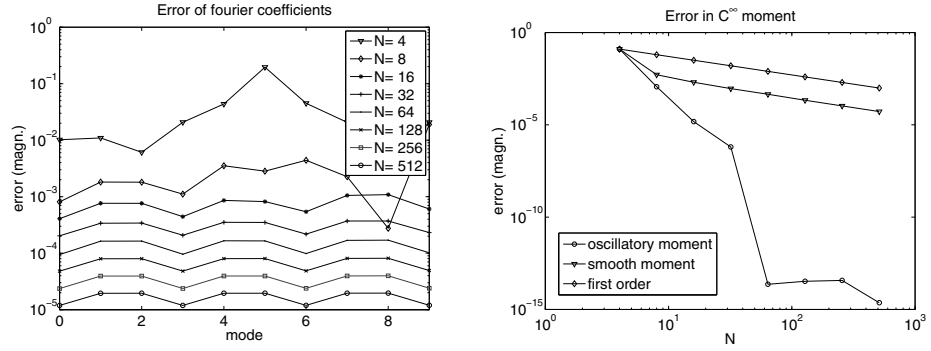

FIG. 2. *Error for $C^\infty(\Omega)$ moments of reaction equation test case in section* 5.2 *at time* $t = 1.0$. *The left plot shows the error of the first 10 Fourier coefficients for various orders. The right plot shows the error for the $C^\infty(\Omega)$ moment $\langle u, \sin\rangle$.*

5.3. Wave equation with discontinuous coefficients. We now consider the wave equation

$$\partial_t u(x,t) + \partial_x \left(\begin{pmatrix} 0 & a(x) \\ b(x) & 0 \end{pmatrix} \cdot u(x,t) \right) = 0 \tag{5.5}$$

on the space-time domain $[0; 2\pi] \times (0; 0.6]$. In contrast to the scalar advection equation, the wave equation test case has characteristics pointing in both spatial directions, i.e., counterpropagating waves. The initial condition is set to

$$u_0(x) = \begin{pmatrix} \cos(4x + 0.2) + 0.1\cos(7x) + 0.001\sin(16x) + 0.0001\sin(44x) \\ \sin(5x + 0.1) + 0.1\sin(9x) + 0.001\cos(20x) + 0.0001\cos(42x) \end{pmatrix} \tag{5.6}$$

and the coefficients are

$$a(x) = b(x) = \begin{cases} 1 & \text{if } x \le 2\pi/3, \\ 0.5 & \text{else.} \end{cases} \tag{5.7}$$

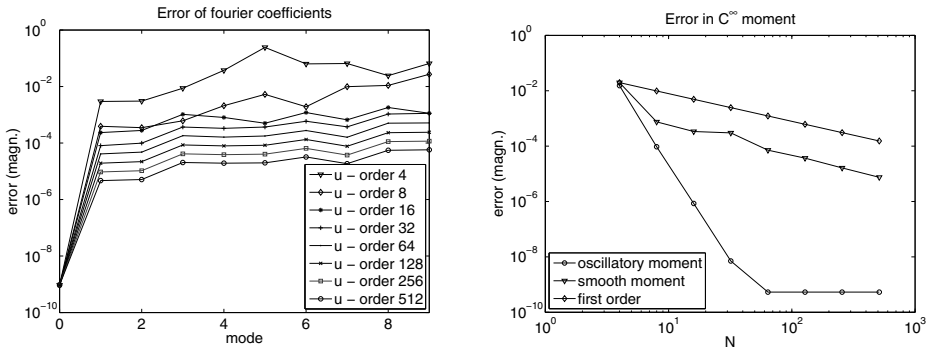

FIG. 3. *Error for $C^\infty(\Omega)$ moments of the wave equation test case in section 5.3 at time $t = 0.6$. The left plot shows the error of the first 10 Fourier coefficients of u_1 for various orders. The right plot shows the moment error convergence for the smooth and oscillatory moment toward the exact value of the moment $\langle u, \sin\rangle$.*

Notice that the two waves of the given system propagate with speed $\sqrt{ab}$ and $-\sqrt{ab}$. By setting a, b as in (5.7) we have two counterpropagating waves through the entire simulation domain.

Figure 3 shows the resulting error plots for the moments of u. The left plot gives the error of the first 10 Fourier coefficients of u_1 (the first component of u). The right plot provides error convergence of the moment for $f = (\sin, \sin)^T$. Both plots are in good agreement with our theoretical predictions, i.e., they provide numerical evidence that our theory also applies to setups with counterpropagating waves. The smooth moment converges slowly with $1/N$, while the oscillatory moment converges exponentially fast as $N \to \infty$.

5.4. Advection equation with discontinuous coefficient and DGFEM. We return to the scalar advection equation $\partial_t u + \partial_x a u = 0$ but discretize the equation in space by the upwind discontinuous Galerkin method. The adjoint problem, to obtain g_N, is solved according to the naturally arising scheme in (3.6). We consider the space-time domain $[0; 2\pi] \times (0; \pi/8]$, which is discretized by six elements and various orders. We use a modal Legendre basis for the discontinuous Galerkin method and apply quadrature rules of sufficient order to integrate all terms exactly. As an initial condition we use

$$u_0(x) = \sin(x) + \sin(2x + 0.32) + \sin(9x + 1.31) + 0.1\sin(12x + 2.11). \tag{5.8}$$

The coefficient $a(x)$ is set to

$$a(x) = \begin{cases} 1.0 & \text{if } 1.0 \le x \le 2\pi/3 + 1.0, \\ 0.5 & \text{else.} \end{cases} \tag{5.9}$$

Notice that the discontinuity does not coincide with one of the internal element boundaries. The resulting error plot for the oscillatory moment with $f(x) = \sin(x)$ is shown in Figure 4. Clearly, exponential error convergence in terms for $N \to \infty$ can be observed, in perfect agreement with the theory of section 3.

5.5. Burgers equation and DGFEM. We consider the inviscid Burgers equation $\partial_t u + \partial_x u^2/2 = 0$ with initial condition

$$u_0(x) = 2 + 0.5\sin 2x \tag{5.10}$$

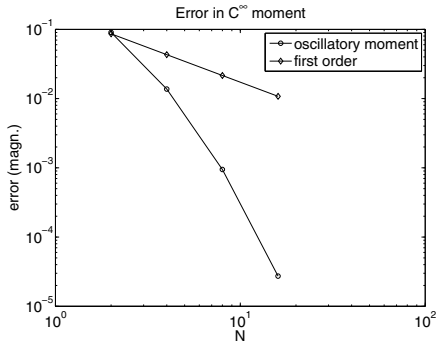

FIG. 4. *Convergence of oscillatory moment* $\langle u_N, g_N \rangle$ *toward* $\langle u, \sin \rangle$ *for DGFEM discretized advection equation test case of section* 5.4. *The domain* Ω *is discretized by six elements. Spectral convergence toward the exact moment can be observed.*

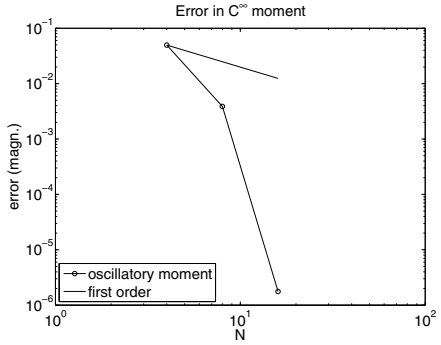

FIG. 5. *Convergence of oscillatory moment* $\langle u_N, g_N \rangle$ *toward* $\langle u, f \rangle$ *for DGFEM discretized Burgers equation test case of section* 5.5. *Spectral convergence toward the exact moment can be observed.*

on the periodic domain $[0; 2\pi]$. We apply a DGFEM discretization with 12 elements and various orders. For this numerical test, we set

$$v_0(x) = \begin{cases} 0 & \text{if } |x - c| > \epsilon, \\ \frac{1}{\epsilon} \exp \frac{-1}{1-((x-c)/\epsilon)^2} & \text{else.} \end{cases} \tag{5.11}$$

We set $c = \pi$, $\epsilon = 0.25$ and advance u, v forward in time until u develops a shock. Once the shock has formed, the solution u cannot be given in a closed formed expression. This makes it impossible to solve the auxiliary problem exactly, neither forward nor backward in time. Instead, we apply Lemma 2.1 and compute the exact moments initially. Notice that the initial condition (5.11) ensures that the lemma is applicable. Figure 5 shows the resulting convergence plot for Burgers equation, confirming that high order information is also available in the nonlinear setup. In Figure 6 the solutions u_N,v_N for $N = 16$ are shown at $t = 0, \pi/4, \pi/2$.

6. Conclusion. In this publication, we considered spectral methods for linear hyperbolic conservation laws with discontinuous coefficients. Even though numerical results suffer from Gibbs oscillations, degrading convergence at first glance, high order information is still contained in the numerical results and can be recovered. This applies to spectral Fourier–Galerkin methods as well as other methods, like the DGFEM,

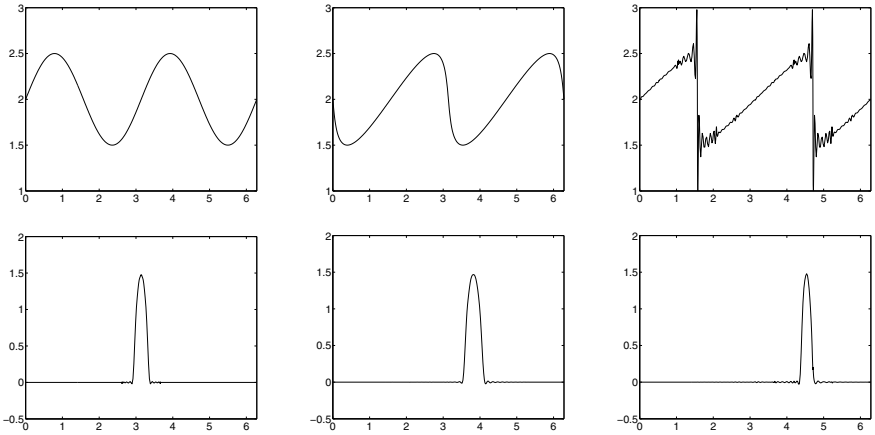

FIG. 6. *Numerical solutions u_N (top) and v_N (bottom) of Burgers equation test case of section 5.5 for 12 elements with $N = 16$ at $t = 0, \pi/4, \pi/2$ (left to right).*

whenever these methods are adjoint-consistent and stable. In both cases information about the moments converges exponentially fast after extraction, including L_2-stable nonlinear problems.

REFERENCES

[1] S. ABARBANEL, D. GOTTLIEB, AND E. TADMOR, *Spectral methods for discontinuous problems*, in Numerical Methods for Fluid Dynamics, II (Reading, 1985), Inst. Math. Appl. Conf. Ser. New Ser. 7, Oxford University Press, New York, 1986, pp. 129–153.

[2] J. P. BOYD, *Chebyshev and Fourier Spectral Methods: Second Revised Edition*, Dover. Mineola, NY, 2013.

[3] C. CANUTO, Y. HUSSAINI, A. QUARTERONI, AND T. A. ZANG, *Spectral Methods: Evolution to Complex Geometries and Applications to Fluid Dynamics*, Sci. Comput., Springer, New York, 2007.

[4] C. CANUTO, Y. HUSSAINI, A. QUARTERONI, AND T. A. ZANG, *Spectral Methods: Fundamentals in Single Domains*, Sci. Comput., Springer, New York, 2007.

[5] Z. CHEN AND C.-W. SHU, *Recovering exponential accuracy in Fourier spectral methods involving piecewise smooth functions with unbounded derivative singularities*, J. Sci. Comput., to appear.

[6] B. COCKBURN, S. HOU, AND C.-W. SHU, *The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws.* IV: *The multidimensional case*, Math. Comp., 54 (1990), pp. 545–581.

[7] B. COCKBURN, S.-Y. LIN, AND C.-W. SHU, *The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws* III: *One-dimensional systems*, J. Comput. Phys., 84 (1989), pp. 90–113.

[8] B. COCKBURN AND C.-W. SHU, *The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws* II: *General framework*, Math. Comput., 52 (1989), pp. 411–435.

[9] B. COCKBURN AND C.-W. SHU, *Runge-Kutta discontinuous Galerkin methods for convection-dominated problems*, J. Sci. Comput., 16 (2001), pp. 173–261.

[10] D. A. DI PIETRO AND A. ERN, *Mathematical Aspects of Discontinuous Galerkin Methods*, Math. Appl., Springer, Berlin, 2011.

[11] W. S. DON, *Numerical study of pseudospectral methods in shock wave applications*, J. Comput. Phys., 110 (1994), pp. 103–111.

[12] B. FORNBERG, *A Practical Guide to Pseudospectral Methods*, Cambridge Monogr. Appl. Comput. Math., Cambridge University Press, Cambridge, UK, 1998.

[13] A. Gelb and J. Tanner, *Robust reprojection methods for the resolution of the Gibbs phenomenon*, Appl. Comput. Harmon. Anal., 20 (2006), pp. 3–25.
[14] D. Gottlieb and J. S. Hesthaven, *Spectral methods for hyperbolic problems*, J. Comput. Appl. Math., 128 (2001), pp. 83–131.
[15] D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications*, CBMS-NSF Regional Conf. Ser. Appl. Math., SIAM, Philadelphia, 1977.
[16] D. Gottlieb, S. A. Orszag, and E. Turkel, *Stability of pseudospectral and finite-difference methods for variable coefficient problems*, Math. Comput., 37 (1981), pp. 293–305.
[17] D. Gottlieb and C.-W. Shu, *On the Gibbs phenomenon* V: *Recovering exponential accuracy from collocation point values of a piecewise analytic function*, Numer. Math., 71 (1995), pp. 511–526.
[18] D. Gottlieb and C.-W. Shu, *On the Gibbs Phenomenon* III: *Recovering exponential accuracy in a sub-interval from a spectral partial sum of a piecewise analytic function*, SIAM J. Numer. Anal., 33 (1996), pp. 280–290.
[19] D. Gottlieb and E. Tadmor, *Recovering Pointwise Values of Discontinuous Data Within Spectral Accuracy*, in Progress and Supercomputing in Computational Fluid Dynamics, E. M. Murman and S. S. Abarbanel, eds., Progr. Sci. Comput. 6, Birkhäuser, Boston, 1985, pp. 357–375.
[20] J. S. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*, Texts in Appl. Math., Springer, New York, 2008.
[21] G. S. Jiang and C.-W. Shu, *On a cell entropy inequality for discontinuous Galerkin methods*, Math. Comp., 62 (1994), pp. 531–531.
[22] P. D. Lax, *Accuracy and resolution in the computation of solutions of linear and nonlinear equations*, in Recent Advances in Numerical Analysis (Proc. Sympos., Math. Res. Center, University of Wisconsin, Madison, WI, 1978), Publ. Math. Res. Center Univ. Wisconsin 41, Academic Press, New York, 1978, pp. 107–117.
[23] Y. Maday, S. M. Ould Kaber, and E. Tadmor, *Legendre pseudospectral viscosity method for nonlinear conservation laws*, SIAM J. Numer. Anal., 30 (1993), pp. 321–342.
[24] W. H. Reed and T. R. Hill, *Triangular Mesh Methods for the Neutron Transport Equation*, Tech. report LA-UR-73-479, Los Alamos Scientific Laboratory, 1973.
[25] C.-W. Shu and P. S. Wong, *A note on the accuracy of spectral method applied to nonlinear conservation laws*, J. Sci. Comput., 10 (1995), pp. 357–369.
[26] E. Tadmor, *Convergence of spectral method for nonlinear conservation laws*, SIAM J. Numer. Anal., 26 (1989), pp. 30–44.
[27] E. Tadmor, *Filters, mollifiers, and the computation of the Gibbs phenomenon*, Acta Numer., 16 (2007), p. 305.
[28] H. Vandeven, *Family of spectral filters for discontinuous problems*, J. Sci. Comput., 6 (1991), pp. 159–192.