Evolution Galerkin methods for hyperbolic systems in two space dimensions

M. Lukáčová - Medviďová [1 2], K.W. Morton[3], and G. Warnecke[1]

Abstract

The subject of the paper is the analysis of three new evolution Galerkin schemes for a system of hyperbolic equations, and particularly for the wave equation system. The aim is to construct methods which take into account all of the infinitely many directions of propagation of bicharacteristics. The main idea of the evolution Galerkin methods is the following: the initial function is evolved using the characteristic cone and then projected onto a finite element space. A numerical comparison is given of the new methods with already existing methods both those based on the use of bicharacteristics as well as commonly used finite difference and finite volume methods. We discuss the stability properties of the schemes and derive error estimates.

Key words: genuinely multidimensional schemes, hyperbolic systems, wave equation, Euler equations, evolution Galerkin schemes

AMS Subject Classification: 35L05, 65M06, 35L45, 35L65, 65M25, 65M15

1 Introduction

It is our belief that the most satisfying methods for approximating evolutionary PDEs are based on approximating the associated evolutionary operator, or its dominant part. Then most advantage seems to be gained by placing this in a Galerkin formulation. This is the basis of the 'modified method of characteristics' pioneered by Douglas and Russell [7], the Lagrange-Galerkin methods of Pironneau [22] and of Benqué *et al.* [1], the Euler–Characteristic–Galerkin methods of Childs and Morton [5] and most recently the methods of Fey [8] — see [16] for a partial review and other references.
For simple problems this approach can be fully exploited. Thus for linear, constant coefficient advection the solution evolves by pure transport along the characteristics. In a Galerkin framework on a uniform mesh all the integrals can be calculated exactly and trivially. Using a continuous piecewise linear approximation in one dimension gives the highly effective third order accurate method first derived by Lesaint [11] and often rediscovered; Childs and Morton [5] showed that the L_2 projection gives the most accurate

[1]Institute of Analysis and Numerics, Otto-von-Guericke-Universität Magdeburg, Universitätsplatz 2, 39 106 Magdeburg, Germany, email: Gerald.Warnecke@mathematik.uni-magdeburg.de

[2]Department of Mathematics, Faculty of Mechanical Engineering, Technical University Brno, Technická 2, 616 39 Brno, Czech Republic, email: Lukacova@fme.vutbr.cz

[3]Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom (also Oxford University Computing Laboratory) email: Bill.Morton@comlab.ox.ac.uk

scheme, but other projections give alternative well-known third order schemes; and tensor product basis functions in more dimensions give tensor product difference operators. The methods are unconditionally stable and can be extended to arbitrary orders of accuracy by either increasing the order of the approximation space or introducing a recovery stage with higher order approximations.

For more complicated problems, and on irregular meshes, approximations and compromises have to be made in the interest of practical efficiency. Thus when the Lagrange-Galerkin method is applied to the Navier–Stokes equations on practical domains there are several approximations in common use [1], [22], [26]: the evolutionary operator is applied to only the convective terms; the trajectories have to be approximated; and all the integrals have to be handled by very careful quadrature. Similarly, when the approach is applied to weather forecasting [10], [25], [27], all the Galerkin integrals are deemed too expensive and a semi-Lagrangian approach is used instead. When the characteristic Galerkin method is applied to scalar nonlinear conservation laws, and the solution may contain shocks, the evolution operator is approximated by using the Brenier transport collapse operator, so that the time step is limited by the resulting inaccuracy [5]. However, the important point in each case is that using an approximate evolution operator over a finite time interval affords the scheme a basic advantage.

Our purpose in this paper is to develop a number of approximate evolution operators for hyperbolic systems of equations in two space dimensions, in particular for the wave equation and later for the Euler equations. Preliminary attempts will then be described to develop these into practical numerical algorithms.

Scalar problems in several space dimensions, or systems that can be reduced to (possibly coupled) scalar equations, pose no essentially new difficulties to evolution–Galerkin methods. But the two-dimensional wave equation, with the solution at a point determined by that over the whole base of the bicharacteristic cone, presents a new challenge which is generic for hyperbolic systems. Thus we concentrate on the development and analysis of methods for that problem. It should be noted here that a corresponding difficulty arises in the solution of the steady subsonic Euler equations where a distinction needs to be made between the elliptic and hyperbolic parts of the system. This has been exploited, for example, in the design of mulitgrid methods by Brandt and Ta'asan [2].

The next section will be devoted to finite difference and evolution Galerkin methods for the wave equation system which are based on the straightforward use of the bicharacteristic cone. We will describe the approach of Butler [3] (and of Prasad [23], [24]) who first used bicharacteristics in order to derive numerical schemes. We will prove a useful lemma which allows us to derive new approximate evolution operators for the wave equation system. Using a projection onto piecewise constants a simple evolution Galerkin method is then derived, called the EG1 scheme.

In the third section we describe a general approach to the use of the bicharacteristics for hyperbolic systems. In the fourth section two alternative evolution Galerkin methods for the wave equation system, the so-called EG2 and EG3 schemes, will be derived. They are based respectively on the Butler approach and on the general theory for hyperbolic problems. As is shown in Section 5, where questions of stability and error estimates are discussed, all of our methods based on using piecewise constant approximations are of first order. The accuracy can be increased by increasing the order of approximation space and the accuracy of the approximate evolution operator. However, we indicate at the end of Section 4 how an alternative finite volume formulation, coupled with a recovery stage, can yield second order accuracy even with the first order approximate evolution operators that have been derived here.

In the last section we present some numerical examples for continuous as well as discontinous initial data. We compare the scheme EG3, closely related to a scheme first derived by Ostkamp [20], [21], with the other two new schemes based on the earlier more ad hoc approach, the EG1 and EG2 schemes. It is shown that the scheme which uses a general theory of bicharacteristics for the hyperbolic systems, viz the EG3 scheme, gives the best numerical results in terms of accuracy. Therefore it will be a good candidate for the construction of higher order schemes [14]. It will be observed that circular symmetry is preserved by the evolution Galerkin schemes very well, which is for example not the case when using finite volume dimensional splitting methods. This is an important property and advantage of the evolution Galerkin schemes, which are also seen to conserve vorticity very well. Moreover, it will be shown that the exact integration in space, used in EG1, EG2, and EG3 schemes, leads to successful resolution also for discontinuous data problems. This does not hold for Butler's and Prasad's schemes, where the upwind bias introduced by the use of the bicharacteristics is undermined by the interpolation procedures based on the cone apex.

2 A straightforward use of the bicharacteristic cone for the wave equation

In a paper that was well ahead of its time in regard to using an approximation to the evolution operator rather than to the differential equation, Butler in 1960 [3] used differences along a set of four bicharacteristics on a square mesh to derive an approximation to the wave equation in two space dimensions. Indeed, he presented the scheme for a general hyperbolic system and applied it very successfully to the unsteady Euler equations in two space dimensions and also to the steady supersonic Euler equations in three dimensions. Considerable attention was given to the use of bicharacteristics by researchers at that time, and his scheme was judged to be particularly effective (see e.g. [6]). Nevertheless, it is little known at the present time and numerical methods, particularly for steady problems, have developed in the interim in quite different directions.

However, both he as well as Prasad and his co-workers who took up the method again in the 1980's - see [23], [24] - lost most of the advantage from the approach by using a poor interpolatory procedure at the old time level. Thus, suppose $\underline{U}_{k\,l}^{n+1}$ is to be predicted from values of the vector $\underline{U}^n$ at the feet of various bicharacteristics at time level n. Butler used biquadratic interpolation over nine points centered at (k, l) for this purpose; so the interpolation used in a given mesh square varied according to the corner of the square that was being updated. The situation is very different with any evolution Galerkin method, where a consistent approximation at one time level is used everywhere in performing the update.

It was Ostkamp in [20], [21] who put some of the ideas of exploiting bicharacteristics into an evolution Galerkin framework. She did so in the context of general hyperbolic systems, and showed how such schemes in the case of the Euler equations linked up with those methods developed by Fey in [8], [9]. It is her work, especially her approximations to the evolution operator, that we shall build on and develop here; but first we consider approximations to the evolution operator based on Butler's approach.

2.1 Two approximations to the evolution operator

In this section we will briefly recall the approach of Butler [3] as well as Prasad and his co-workers [23], [24] to deriving finite difference methods for the wave equation. For later generalisation, particularly to the Euler equations, it is important to write the wave equation as a first order system in the following form

$$\begin{aligned} \phi_t + c(u_x + v_y) &= 0 \\ u_t + c\phi_x &= 0 \\ v_t + c\phi_y &= 0 \end{aligned} \tag{2.1}$$

using the unknown functions ϕ, u, v. Then any smooth solution ϕ satisfies the second order wave equation, but the system (2.1) also admits solutions with nonzero, but constant, vorticity. Consider a characteristic cone corresponding to system (2.1) with the apex $P = (x, y, t + \Delta t)$ and the base points $Q(\theta) = (x + c\Delta t \cos\theta, y + c\Delta t \sin\theta, t)$ parametrized by the angle $\theta \in [0, 2\pi]$, see Figure 1. The lines from $Q(\theta)$ to P generating the mantle of the cone are bicharacteristics. Differentiation in the direction of these bicharacteristics is given by

$$\frac{\mathrm{d}}{\mathrm{d}\sigma} \equiv \frac{\partial}{\partial t} - c\cos\theta\frac{\partial}{\partial x} - c\sin\theta\frac{\partial}{\partial y}. \tag{2.2}$$

Hence by combining the three equations of (2.1) with weights $(1, -\cos\theta, -\sin\theta)$ we obtain the equation

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}(\phi - u\cos\theta - v\sin\theta) = -S \tag{2.3}$$

with the source term

$$S(\tilde{t}, \theta) = c\left[u_x(\tilde{x}, \tilde{y}, \tilde{t})\sin^2\theta - (u_y(\tilde{x}, \tilde{y}, \tilde{t}) + v_x(\tilde{x}, \tilde{y}, \tilde{t}))\sin\theta\cos\theta + v_y(\tilde{x}, \tilde{y}, \tilde{t})\cos^2\theta\right], \tag{2.4}$$

where $(\tilde{x}, \tilde{y}) = (x + c(t + \Delta t - \tilde{t})\cos\theta, y + c(t + \Delta t - \tilde{t})\sin\theta)$ and $\tilde{t} \in [t, t + \Delta t]$. Integrating (2.3) along the bicharacteristic, from the point $Q(\theta)$ on the base of the bicharacteristic cone to its apex P, gives

$$[\phi]_Q^P - [u]_Q^P\cos\theta - [v]_Q^P\sin\theta = -\int_Q^P S(\tilde{t}, \theta)\mathrm{d}\tilde{t}. \tag{2.5}$$

Now integrating this around the cone, i.e. over θ, gives the integral representation

$$\begin{aligned} \phi_P = &\frac{1}{2\pi}\int_0^{2\pi}[\phi_Q - u_Q\cos\theta - v_Q\sin\theta]\mathrm{d}\theta \\ &- \frac{1}{2\pi}\int_t^{t+\Delta t}\int_0^{2\pi} S(\tilde{t}, \theta)\mathrm{d}\theta\mathrm{d}\tilde{t}. \end{aligned} \tag{2.6}$$

Similarly, by increasing the weights multiplying (2.1) by a factor $\cos\theta$ or $\sin\theta$, one obtains

$$\begin{aligned} u_P = &\frac{1}{\pi}\int_0^{2\pi}[-\phi_Q\cos\theta + u_Q\cos^2\theta + v_Q\sin\theta\cos\theta]\mathrm{d}\theta \\ &+ \frac{1}{\pi}\int_t^{t+\Delta t}\int_0^{2\pi} S(\tilde{t}, \theta)\cos\theta\mathrm{d}\theta\mathrm{d}\tilde{t}, \end{aligned} \tag{2.7}$$

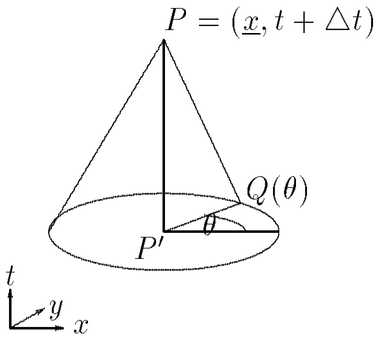

Figure 1: Bicharacteristic along the Mach cone through P and $Q(\underline{\theta})$

$$
\begin{aligned}
v_P = & \frac{1}{\pi} \int_0^{2\pi} [-\phi_Q \sin\theta + u_Q \sin\theta\cos\theta + v_Q \sin^2\theta] \mathrm{d}\theta \\
& + \frac{1}{\pi} \int_t^{t+\Delta t} \int_0^{2\pi} S(\tilde{t}, \theta) \sin\theta \mathrm{d}\theta \mathrm{d}\tilde{t}.
\end{aligned}
\tag{2.8}
$$

Let us note that we have obtained an integral representation of the exact evolution operator. This will be the basis for numerical approximations. The integral of S over intermediate time levels is the term most needing attention. We shall use quadrature in time to construct approximate evolution operators. To neglect S entirely can easily be shown to give an inconsistent approximation. More commonly, the rectangle rule or the trapezoidal rule can be applied to give an $O(\Delta t)$ or $O(\Delta t^2)$ approximation, respectively. The rectangle rule applied to (2.6) gives simply

$$
\phi_P = \frac{1}{2\pi} \int_0^{2\pi} [\phi_Q - u_Q \cos\theta - v_Q \sin\theta] \mathrm{d}\theta - \frac{\Delta t}{2\pi} \int_0^{2\pi} S(t, \theta) \mathrm{d}\theta + O(\Delta t^2), \tag{2.9}
$$

with corresponding formulas for u_P and v_P.
On the other hand, Butler [3] used the trapezoidal rule to obtain the following approximation

$$
\begin{aligned}
\frac{1}{2\pi} \int_t^{t+\Delta t} \int_0^{2\pi} S(\tilde{t}, \theta) \mathrm{d}\theta \mathrm{d}\tilde{t} &= \tfrac{1}{2}\Delta t \Big[\frac{1}{2\pi} \int_0^{2\pi} S(t + \Delta t, \theta) \mathrm{d}\theta + \frac{1}{2\pi} \int_0^{2\pi} S(t, \theta) \mathrm{d}\theta \Big] + O(\Delta t^3) \\
&= \tfrac{1}{2}\Delta t \Big[\tfrac{1}{2} c (u_x + v_y)_P + \frac{1}{2\pi} \int_0^{2\pi} S(t, \theta) \mathrm{d}\theta \Big] + O(\Delta t^3).
\end{aligned}
\tag{2.10}
$$

However, this has introduced the unknown derivatives u_x and v_y at the point P for which we want to solve. Butler solved this problem by eliminating these derivatives via integration of the equations along a time-like bicharacteristic curve $P'P$, where $P' \equiv (x, y, t)$ and $P \equiv (x, y, t + \Delta t)$ (see Figure 1). Thus, using the same trapezoidal rule to integrate the first equation of (2.1) from P' to P gives

$$
\phi_P - \phi_{P'} + \tfrac{1}{2} c \Delta t \left[(u_x + v_y)_P + (u_x + v_y)_{P'} \right] = O(\Delta t^3). \tag{2.11}
$$

Eliminating the derivatives at P from equations (2.10) and (2.11) we get

$$
\begin{aligned}
\phi_P = & \frac{1}{\pi} \int_0^{2\pi} [\phi_Q - u_Q \cos\theta - v_Q \sin\theta] \mathrm{d}\theta - \phi_{P'} \\
& + \frac{c\Delta t}{2} (u_x + v_y)_{P'} - \frac{\Delta t}{2\pi} \int_0^{2\pi} S(t, \theta) \mathrm{d}\theta + O(\Delta t^3).
\end{aligned}
\tag{2.12}
$$

Similar expressions follow from (2.7) and (2.8) for u and v, respectively,

$$u_P = \frac{1}{\pi}\int_0^{2\pi}[-\phi_Q\cos\theta + u_Q\cos^2\theta + v_Q\sin\theta\cos\theta]\mathrm{d}\theta + \frac{\Delta t}{2\pi}\int_0^{2\pi} S(t,\theta)\cos\theta\mathrm{d}\theta + O(\Delta t^3), \tag{2.13}$$

$$v_P = \frac{1}{\pi}\int_0^{2\pi}[-\phi_Q\sin\theta + u_Q\sin\theta\cos\theta + v_Q\sin^2\theta]\mathrm{d}\theta + \frac{\Delta t}{2\pi}\int_0^{2\pi} S(t,\theta)\sin\theta\mathrm{d}\theta + O(\Delta t^3). \tag{2.14}$$

These equations give explicitly the values of ϕ, u and v at the point P, i.e. at time $t+\Delta t$, in terms of the values at t. The error in equations (2.12) - (2.14) is $O(\Delta t^3)$ in one time step, so we obtain second order schemes with respect to time; similarly, (2.9) gives an $O(\Delta t)$ approximation.
Now on the basis of these equations a sequence of finite difference schemes can be derived. We can replace integration with respect to θ by a suitable numerical quadrature and use an appropriate interpolation formula to determine the intermediate values of ϕ, u, v and their derivatives at the points Q on the base of the bicharacteristic cone in terms of the mesh points.
Butler [3], for example, used four bicharacteristics to obtain a finite difference scheme. In fact he approximated the line integral with respect to θ by the trapezoidal rule and took the values on the base of the bicharacteristic cone for $\theta = 0, \frac{\pi}{2}, \pi$ and $\frac{3\pi}{2}$. The finite difference scheme obtained in this manner is given in the Appendix.
On the other hand, Prasad and his co-workers used Simpson's rule with 8 or 16 subdivisions of the interval $[0, 2\pi]$ but, as one of the referees pointed out to us, this choice fails to recognize the special suitability of the trapezoidal rule for integrating a periodic function. Our numerical experiments, described in Section 6, have indeed shown that these schemes have no advantages over those of Butler.

2.2 A useful lemma

We observe that the source S is composed of tangential derivatives on the surface of the bicharacteristic cone so that its integral can be simplified through integration by parts; and for later use we state a more general result as a lemma.

Lemma 2.1 *Suppose $w \in C^1(\mathbb{R}^2)$, and $p \in C^1(\mathbb{R})$ is 2π-periodic. Then integrating round the circle of radius a, with a general point denoted by $Q \equiv (a\cos\theta, a\sin\theta)$, gives*

$$\int_0^{2\pi} p'(\theta)w(Q)\mathrm{d}\theta - a\int_0^{2\pi} p(\theta)[w_x(Q)\sin\theta - w_y(Q)\cos\theta]\mathrm{d}\theta = 0. \tag{2.15}$$

Proof: Consider the integral of $\frac{\mathrm{d}}{\mathrm{d}\theta}[p(\theta)w(Q)]$, noting $\frac{\mathrm{d}}{\mathrm{d}\theta} = -a(\sin\theta\frac{\partial}{\partial x} - \cos\theta\frac{\partial}{\partial y})$. □

Taking $p = \sin\theta, w = u$ and $p = -\cos\theta, w = v$ with $a = c\Delta t$ gives from the definition of S in (2.4)

$$\Delta t\int_0^{2\pi} S(t,\theta)\mathrm{d}\theta = \int_0^{2\pi}[u_Q\cos\theta + v_Q\sin\theta]\mathrm{d}\theta. \tag{2.16}$$

The importance of eliminating the derivatives in the definition of S, to give the right-hand side of (2.16), is that this formula can be applied to piecewise constant approximations, as we shall do in the next section.

Approximate evolution operator for EG1

Thus, from the rectangle rule (2.9) and (2.16) we have the following approximation of (2.6)

$$\phi_P = \frac{1}{2\pi}\int_0^{2\pi}[\phi_Q - 2u_Q\cos\theta - 2v_Q\sin\theta]\mathrm{d}\theta + O(\Delta t^2). \tag{2.17}$$

Formulae similar to (2.16) can be derived for the integrals of $S\sin\theta$ and $S\cos\theta$, so as to obtain corresponding approximations to u_P and v_P from the rectangle rule applied to (2.7) and (2.8), namely

$$u_P = \frac{1}{\pi}\int_0^{2\pi}[-\phi_Q\cos\theta + u_Q(3\cos^2\theta - 1) + 3v_Q\sin\theta\cos\theta]\mathrm{d}\theta + O(\Delta t^2), \tag{2.18}$$

$$v_P = \frac{1}{\pi}\int_0^{2\pi}[-\phi_Q\sin\theta + 3u_Q\sin\theta\cos\theta + v_Q(3\sin^2\theta - 1)]\mathrm{d}\theta + O(\Delta t^2). \tag{2.19}$$

Equations (2.17), (2.18) and (2.19) define an approximate evolution operator for the wave equation system. It gives a first order approximation with respect to the time. In the following we show how this leads to a scheme to be called EG1.

2.3 An evolution Galerkin scheme

For the general hyperbolic system that we shall consider later we denote by $E(s) : (H^k(I\!\!R^d))^m \to (H^k(I\!\!R^d))^m$ the exact evolution operator, acting on the Sobolev spaces $H^k(I\!\!R^d)$, associated with a time step s for the system, i.e.

$$\underline{U}(\cdot, t+s) = E(s)\underline{U}(\cdot, t). \tag{2.20}$$

We suppose that S_h is a finite element space consisting of piecewise polynomials of order r. Let $\underline{U}^n$ be an approximation in the space S_h to the exact solution $\underline{U}(\cdot, t_n)$ at a time $t_n > 0$ and take $E_\triangle : S_h \to (H^k(I\!\!R^d))^m$ to be a suitable approximation to the exact evolution operator $E(\Delta t)$. We denote by $P_h : (H^k(I\!\!R^d))^m \to S_h$ the L^2-projection onto S_h. Then we can define an evolution Galerkin method.

Definition 2.1 Starting from some initial value $\underline{U}^0 \in S_h$ at time $t = 0$, the *evolution Galerkin method (EG)* is recursively defined by means of

$$\underline{U}^{n+1} = P_h E_\triangle \underline{U}^n. \tag{2.21}$$

For simplicity we assume constant time steps Δt, i.e. $t_n = n\Delta t$. The method is uniquely determined by the approximate evolution operator $E_\triangle$ and the projection P_h.

In the present paper, in order to relate our schemes to the difference schemes given by Butler [3], Prasad *et al.* [24] and also to link up with the well known Lagrange–Galerkin methods for the linear advection scheme (see Morton [17] for a summary of these), we shall limit our consideration to cases where the space S_h is composed of piecewise constant functions on a square mesh. The resulting schemes will therefore be only first order schemes, even when E_Δ is approximated to second order. Higher order accuracy can be obtained either by using higher degree polynomials in S_h, or by inserting a recovery step R_h before the evolution step in (2.21) to give $\underline{U}^{n+1} = P_h E_\Delta R_h \underline{U}^n$, see [16], [17], or [18] for more details. An indication of how this may also be achieved instead through a finite volume formulation will be given in Section 4.4.

Now let Ω be our computational domain. We construct a mesh for Ω, which consists of the square mesh cells $\Omega_{kl} \equiv [(k-\frac{1}{2})h, (k+\frac{1}{2})h] \times [(l-\frac{1}{2})h, (l+\frac{1}{2})h]$, where $k, l \in \mathbb{Z}$, and $h > 0$ is the mesh size parameter. In what follows we will work with the L^2-projection given by integral averages onto a space S_h of piecewise constant step functions,

$$P_h \underline{U} = \sum_{k,l \in \mathbb{Z}} \left(\frac{1}{h^2} \int_{\Omega_{kl}} \underline{U}(x,y) dx dy \right) \chi_{kl} \qquad \underline{U} \in (L^2(\mathbb{R}^d))^m, \tag{2.22}$$

where χ_{kl} is the characteristic function for the square mesh cell Ω_{kl}. Now, the evolution Galerkin algorithm using (2.17) takes the form

$$\phi_{kl}^{n+1} = \frac{1}{2\pi\, h^2} \int_{\Omega_{kl}} \int_0^{2\pi} \left[\phi_Q^n - 2u_Q^n \cos\theta - 2v_Q^n \sin\theta \right] \mathrm{d}\theta \mathrm{d}x \mathrm{d}y, \tag{2.23}$$

where as before $Q \equiv (x + c\Delta t \cos\theta, y + c\Delta t \sin\theta)$. Similar expressions follow from (2.18), (2.19) for u and v. We emphasize that the integrals in (2.23) are not approximated using numerical quadrature as in the Butler and Prasad schemes. We compute these triple integrals exactly using the fact that the approximate functions ϕ^n, u^n, v^n are piecewise constant. The rather technical computations of these integrals with respect to θ, x and y have been carried out for $c\Delta t \leq h$. Using standard finite difference notation for the first and second order central differences,

$$\begin{aligned} \Delta_{0x} f(x) &:= 1/2[f(x+h) - f(x-h)] \\ \delta_x^2 f(x) &:= f(x+h) - 2f(x) + f(x-h), \end{aligned} \tag{2.24}$$

the resulting formula for ϕ can be written in the following form

$$\begin{aligned} \phi^{n+1} = \left[1 + \frac{\nu}{\pi}(\delta_x^2 + \delta_y^2) + \frac{\nu^2}{4\pi} \delta_x^2 \delta_y^2 \right] \phi^n \; - \nu(1 + \frac{2\nu}{3\pi}\delta_y^2)\Delta_{0x} u^n \\ - \nu(1 + \frac{2\nu}{3\pi}\delta_x^2)\Delta_{0y} v^n, \end{aligned} \tag{2.25}$$

where we denote by ν the CFL number $c\Delta t/h$. When (2.25) is coupled with the corresponding formulae for u^{n+1} and v^{n+1}, which are given in the Appendix, we call this evolution Galerkin scheme for the wave equation EG1. Note the unusual factor of π in the coefficients of this scheme, which comes from the exact computation of the integrals in the approximate evolution operator. This is a novel feature of all of our EG schemes that seems to reflect the geometry of the characteristic cone.

It is interesting to compare this scheme with other more familiar schemes in the particular case that the data and solution do not vary in the y - direction. Then the wave equation reduces to the advection of $\phi + cu$ and $\phi - cu$ along the characteristics $\mathrm{d}x/\mathrm{d}t = c$ and $\mathrm{d}x/\mathrm{d}t = -c$, respectively; so the evolution operator can be evaluated exactly. The result, for piecewise constant approximations, is the first order upwind scheme for each of $\phi \pm cu$, so that in combination we obtain

$$\phi^{n+1} = (1 + \frac{\nu}{2}\delta_x^2)\phi^n - \nu \Delta_{0x} u^n, \tag{2.26}$$

with a corresponding formula for u^{n+1}. For comparison, (2.25) reduces to a similar scheme but with the diffusion coefficient of $\nu/2$ reduced to ν/π in the ϕ update and increased to $2\nu/\pi$ in the u update. On the other hand, the Lax-Wendroff scheme, to which the Butler scheme would reduce in this case, is similar but with a diffusion coefficient of $\nu^2/2$ in both cases. One can show that it is the approximation of the evolution operator by the rectangle rule that is responsible for the reduction of the damping in (2.25) compared with (2.26), but (2.25) can still be regarded as an upwind scheme and it is more damped than Lax-Wendroff for $\nu < 2/\pi$.

3 General hyperbolic systems

Now we describe a general approach to the derivation of the exact evolution operator for any first order hyperbolic system. The role of the bicharacteristics will be illustrated more precisely. We consider a general hyperbolic system in d space dimensions

$$\underline{U}_t + \sum_{k=1}^{d} \underline{\underline{A}}_k \underline{U}_{x_k} = 0, \;\; \underline{x} = (x_1, \ldots, x_d)^T \in I\!R^d, \tag{3.1}$$

where the coefficient matrices $\underline{\underline{A}}_k$, $k = 1, \ldots, d$ are elements of $I\!R^{m\times m}$ and the dependent variables are $\underline{U} = (u_1, \ldots, u_m)^T \in I\!R^m$. Because of the assumed hyperbolicity of the system we have m real eigenvalues λ_j, $j = 1, \ldots, m$ and corresponding linearly independent right eigenvectors $\underline{r}_j = \underline{r}_j(\underline{n})$, $j = 1, \ldots, m$ of the matrix pencil $\underline{\underline{A}}(\underline{n}) := \sum_{k=1}^{d} n_k \underline{\underline{A}}_k$ for any unit vector $\underline{n} = (n_1, \ldots, n_d)^T \in I\!R^d$. Since a common factor is irrelevant we assume $|\underline{n}| = 1$. In the case $d = 2$ we have replaced $\underline{n}$ on the unit circle by the parameter θ in Section 2.
We denote by $\underline{\underline{R}} = \underline{\underline{R}}(\underline{n}) := (\underline{r}_1, \ldots, \underline{r}_m)$ the matrix of the right column eigenvectors. For any direction $\underline{n}$ the characteristic variables $\underline{W} = \underline{W}(\underline{n}) = (w_1, \ldots, w_m)^T$ for a general, possibly nonlinear, hyperbolic system, are defined by $\partial \underline{W}(\underline{n}) = \underline{\underline{R}}^{-1}(\underline{n})\partial \underline{U}$, i.e. for constant coefficient matrices this can be integrated to yield

$$\underline{W} = \underline{\underline{R}}^{-1}\underline{U}, \;\; \underline{U} = \underline{\underline{R}}\,\underline{W}.$$

Multiplying (3.1) by $\underline{\underline{R}}^{-1}$ from the left we obtain the characteristic system

$$\underline{W}_t + \sum_{k=1}^{d} \underline{\underline{B}}_k \underline{W}_{x_k} = 0 \tag{3.2}$$

where $\underline{\underline{B}}_k := \underline{\underline{R}}^{-1}\underline{\underline{A}}_k\underline{\underline{R}} = (b_{ij}^k)_{i,j=1}^m$. We introduce the decomposition $\underline{\underline{B}}_k = \underline{\underline{\Lambda}}_k + \underline{\underline{B}}'_k$, where $\underline{\underline{\Lambda}}_k$ is the matrix containing the diagonal part of $\underline{\underline{B}}_k$. This gives a quasi-diagonalised system

$$\underline{W}_t + \sum_{k=1}^{d} \underline{\underline{\Lambda}}_k \underline{W}_{x_k} = -\sum_{k=1}^{d} \underline{\underline{B}}'_k \underline{W}_{x_k} =: \underline{S}, \tag{3.3}$$

which reduces to the diagonal case, i.e. $\underline{S} = \underline{0}$, in the special case that all matrices $\underline{\underline{A}}_k$ commute. The j-th bicharacteristic corresponding to the j-th equation of the system (3.3) is defined by

$$\frac{\mathrm{d}\underline{x}_j}{\mathrm{d}t} = \underline{b}_{jj}(\underline{n}) := (b_{jj}^1, \ldots, b_{jj}^d)^T. \tag{3.4}$$

We integrate the j-th equation of the system (3.3) from the point P down to the point $Q_j(\underline{n})$, where the bicharacteristic hits the plane through P' (see Figure 1). For a linear constant coefficient problem this will be a straight line. In this case $\underline{\underline{A}}(\underline{n})$ is constant and we assume this in what follows. For nonlinear systems this is an approximation which will place a limit on the choice of Δt.
Integration along the bicharacteristics introduces a formula for the characteristic variables

$$w_j(P, \underline{n}) - w_j(Q_j(\underline{n}), \underline{n}) = S'_j(\underline{n}), \ \ j = 1, \ldots, m, \tag{3.5}$$

with $S'_j(\underline{n}) = \int_t^{t+\Delta t} S_j(\underline{x}_j(\tilde{t}, \underline{n}), \tilde{t}, \underline{n}) \, \mathrm{d}\tilde{t}$. Multiplication of (3.5) by $\underline{\underline{R}}$ from the left and integration of the variable $\underline{n}$ over the unit sphere O in $\mathbb{R}^d$ (the base of the Mach cone) leads to an integral representation

$$\begin{aligned} \underline{U}(P) = \underline{U}(\underline{x}, t + \triangle t) &= \frac{1}{|O|} \int_O \underline{\underline{R}}(\underline{n}) \begin{bmatrix} w_1(Q_1(\underline{n}), \underline{n}) \\ \vdots \\ w_m(Q_m(\underline{n}), \underline{n}) \end{bmatrix} \mathrm{d}O + \underline{\tilde{S}} \\ &= \frac{1}{|O|} \int_O \sum_{j=1}^{m} w_j(Q_j(\underline{n}), \underline{n}) \, \underline{r}_j(\underline{n}) \, \mathrm{d}O + \underline{\tilde{S}} \end{aligned} \tag{3.6}$$

with

$$\underline{\tilde{S}} = (\tilde{S}_1, \ldots, \tilde{S}_m)^T := \frac{1}{|O|} \int_O \underline{\underline{R}}(\underline{n}) \, \underline{S}'(\underline{n}) \, \mathrm{d}O = \frac{1}{|O|} \int_O \int_t^{t+\Delta t} \underline{\underline{R}}(\underline{n}) \, \underline{S}(\tilde{t}, \underline{n}) \, \mathrm{d}\tilde{t} \, \mathrm{d}O.$$

This is an exact representation formula for the evolution operator. The second term contains the integral between the two time levels t and $t + \Delta t$ which in general cannot be evaluated exactly. Thus suitable approximations for the source term $\underline{\tilde{S}}$ may be used. The main purpose of this paper is to explore various possibilities as a starting point for the development of numerical schemes.

4 Further evolution Galerkin methods for the wave equation system

We will now deal more fully with the wave equation system (2.1). First we derive an alternative form for the awkward source term in (2.6) - (2.8). Then we apply Butler's ideas as well as the general theory for linear hyperbolic systems described in Section 3 to the wave equation system and obtain two further approximate evolution operators.

4.1 Exact evolution operator for the wave equation

In order to compare with other methods of approximation, it is useful to transform the exact evolution operators of (2.6) - (2.8) by applying Lemma 2.1 at each time level, and also to eliminate the angle θ. Thus from (2.15), in the same way that (2.16) was obtained, we have

$$\begin{aligned} &\frac{1}{2\pi} \int_t^{t+\Delta t} \int_0^{2\pi} S(\tilde{t}, \theta) \mathrm{d}\theta \mathrm{d}\tilde{t} = \int_0^{\Delta t} \frac{1}{2\pi\tau} \int_0^{2\pi} [u \cos\theta + v \sin\theta] \Big|_{\tilde{t} = t + \Delta t - \tau} \mathrm{d}\theta \, \mathrm{d}\tau \\ &= \int_0^{\Delta t} \frac{1}{2\pi c \tau^2} \int_{D_\tau} \mathrm{div}(u, v) \, \mathrm{d}x \mathrm{d}y \mathrm{d}\tau = \frac{c}{2} \int_0^{\Delta t} \langle \mathrm{div}(u, v) \rangle_\tau \mathrm{d}\tau, \end{aligned} \tag{4.1}$$

where $\langle\cdot\rangle_\tau$ represents an average over the disk D_τ, of radius $c\tau$, given by the intersection of the characteristic cone with the plane at time level $\tilde{t}$. If we also denote by $(\cdot)_{\Delta t}$ an average over the perimeter of the characteristic cone base, i.e. the circle of radius $c\Delta t$, the exact evolution operator (2.6) defining ϕ_P can be written in the following form

$$\phi_P = (\phi)_{\Delta t} - \frac{1}{2}c\left[\Delta t\langle\mathrm{div}(u,v)\rangle_{\Delta t} + \int_0^{\Delta t}\langle\mathrm{div}(u,v)\rangle_\tau \mathrm{d}\tau\right]. \tag{4.2}$$

The relationship with the first equation of (2.1) is now much more obvious.

Obtaining similar formulae for u and v is more difficult, because of the extra factors $\cos\theta$ and $\sin\theta$; but approximations to any order of Δt are easily derived from combining quadrature formulae in the t-direction with Taylor expansions in the (x,y) - plane. Thus, for example, an expansion about the axis points $(\underline{x},\tilde{t})$ leads to the update for u

$$\begin{aligned} u_P = \left[u + \frac{1}{8}(c\Delta t)^2(3u_{xx} + u_{yy} + 2v_{xy})\right]_t - c\Delta t\left[\phi_x + \frac{1}{8}(c\Delta t)^2\nabla^2\phi_x\right]_t \\ + \frac{1}{4}c^2\int_0^{\Delta t}\tau\left[u_{xx} - u_{yy} + 2v_{xy}\right]_{\tilde{t}}\mathrm{d}\tau + O(\Delta t^4), \end{aligned} \tag{4.3}$$

with a similar formula for v. Such expansions are useful in assessing the truncation errors of schemes such as those derived below, since they allow the error committed in approximating the evolution operator to be separated from the projection error. They may also be compared with the Taylor expansion in the t - direction, followed by substitution from the differential equation, that is used in the development of Lax-Wendroff difference schemes.

4.2 EG2, a Butler-based scheme

We reconsider the approach of Butler outlined in Section 2.1, which uses the trapezoidal rule in time to obtain an $O(\Delta t^2)$ appoximation to the evolution operator, but we eliminate $(u_x + v_y)$ in (2.11) and (2.12) at P' as well as P. From Taylor expansions as used in obtaining (4.3), followed by application of Gauss' theorem, we obtain

$$\begin{aligned} \pi(c\Delta t)^2(u_x+v_y)_{P'} &= \int_O (u_x+v_y)\,\mathrm{d}x\mathrm{d}y + O(\Delta t^4) = \oint[u\mathrm{d}y - v\mathrm{d}x] + O(\Delta t^4) \\ &= c\Delta t\int_0^{2\pi}[u_Q\cos\theta + v_Q\sin\theta]\mathrm{d}\theta + O(\Delta t^4). \end{aligned} \tag{4.4}$$

Approximate evolution operator for EG2

Combining (4.4) with (2.16) we see that the last two terms of (2.12) cancel and we obtain

$$\phi_P = \frac{1}{\pi}\int_0^{2\pi}[\phi_Q - u_Q\cos\theta - v_Q\sin\theta]\mathrm{d}\theta - \phi_{P'} + O(\Delta t^3). \tag{4.5}$$

In obtaining similar approximations for u_P and v_P we note that the integrals of $S\cos\theta$ and $S\sin\theta$ at $t+\Delta t$ give zero contributions; and at t we apply Lemma 2.1 first with $p = \sin\theta\cos\theta, w = u$ and then with $p = -\cos^2\theta, w = v$ to obtain

$$u_P = \frac{1}{\pi}\int_0^{2\pi}[-\phi_Q\cos\theta + u_Q(2\cos^2\theta - \tfrac{1}{2}) + 2v_Q\sin\theta\cos\theta]\mathrm{d}\theta + O(\Delta t^3), \tag{4.6}$$

and similarly we have

$$v_P = \frac{1}{\pi}\int_0^{2\pi}[-\phi_Q \sin\theta + 2u_Q \sin\theta\cos\theta + v_Q(2\sin^2\theta - \tfrac{1}{2})]\mathrm{d}\theta + O(\Delta t^3). \qquad (4.7)$$

Equations (4.5), (4.6) and (4.7) define a new approximate evolution operator for the wave equation system. It gives a second order approximation with respect to the time. Using the projection onto a space of piecewise constant functions we will obtain a numerical scheme which will be referred to as the EG2 scheme, but this is now only first order accurate. The finite difference formulation of this scheme can be found in the Appendix.

4.3 EG3, a scheme derived from the general theory

Our aim will be to apply the theory for a general linear hyperbolic system and following Ostkamp [20], [21] to derive another scheme for the wave equation system. System (2.1) in two space dimensions can be written down in the following form

$$\underline{U}_t + \underline{\underline{A}}_1\underline{U}_x + \underline{\underline{A}}_2\underline{U}_y = 0, \quad \underline{x} = (x,y)^T \in I\!R^2, \qquad (4.8)$$

where the, noncommuting, coefficient matrices $\underline{\underline{A}}_1, \underline{\underline{A}}_2 \in I\!R^{3\times 3}$ are defined by

$$\underline{\underline{A}}_1 := \begin{pmatrix} 0 & c & 0 \\ c & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \underline{\underline{A}}_2 := \begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ c & 0 & 0 \end{pmatrix}.$$

Here $c \in I\!R$ denotes the speed of sound and $\underline{U} = (\phi, u, v)^T \in I\!R^3$ is the vector of dependent variables. We have three eigenvalues $\lambda_1 = -c, \lambda_2 = 0, \lambda_3 = c$ and corresponding linear independent right eigenvectors

$$\underline{r}_1 = \begin{bmatrix} -1 \\ \cos\theta \\ \sin\theta \end{bmatrix}, \quad \underline{r}_2 = \begin{bmatrix} 0 \\ \sin\theta \\ -\cos\theta \end{bmatrix}, \quad \underline{r}_3 = \begin{bmatrix} 1 \\ \cos\theta \\ \sin\theta \end{bmatrix}$$

of the matrix pencil $\underline{\underline{A}}(\underline{n}) := \underline{\underline{A}}_1\cos\theta + \underline{\underline{A}}_2\sin\theta$ for any unit vector $\underline{n} = (n_x, n_y)^T = (\cos\theta, \sin\theta)^T \in I\!R^2$.

As described in Section 3 we multiply system (4.8) by $\underline{\underline{R}}^{-1} = (\underline{r}_1, \underline{r}_2, \underline{r}_3)^{-1}$ and obtain the characteristic system (3.2) for the wave equation. The characteristic variables $\underline{W}$ are:

$$\underline{W}(\underline{n}) = \underline{\underline{R}}^{-1}(\underline{n})\underline{U} = \begin{bmatrix} \frac{1}{2}(-\phi + u\cos\theta + v\sin\theta) \\ u\sin\theta - v\cos\theta \\ \frac{1}{2}(\phi + u\cos\theta + v\sin\theta) \end{bmatrix}. \qquad (4.9)$$

It is important to point out that in this approach we multiply the system for the wave equation with the whole matrix of all three right eigenvectors $\underline{\underline{R}}^{-1}$, whereas in the previous approach, i.e. (2.6) – (2.8), only the first eigenvector $\underline{r}_1$ was taken into account. Although further considerations will be more or less similar to those made before, this is the crucial point that leads to new representation formulae for the exact evolution operator.

Let us denote the footpoints by

$$Q_1(\theta) = (\underline{x} + c\triangle t\underline{n}(\theta), t), \quad Q_2 = (\underline{x}, t), \quad Q_3(\theta) = (\underline{x} - c\triangle t\underline{n}(\theta), t)$$

and the points on the bicharacteristic curve Q_2P by

$$\tilde{Q}_2 = (\underline{x}, \tilde{t}), \qquad \tilde{t} \in [t, t+\Delta t].$$

The representation formulae analogous to equation (3.6) (see [20], pp. 50-53) are, after a short computation,

$$\begin{aligned}
\phi(\underline{x}, t+\triangle t) &= \frac{1}{2\pi}\int_0^{2\pi}[\phi(Q_1(\theta)) - u(Q_1(\theta))\,\cos\theta - v(Q_1(\theta))\,\sin\theta]\,\mathrm{d}\theta \\
&\quad -\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\triangle t} S(\underline{x}+c\tau\underline{n}(\theta),\tilde{t},\theta)\,\mathrm{d}\tau\,\mathrm{d}\theta, \qquad (4.10)
\end{aligned}$$

$$\begin{aligned}
u(\underline{x}, t+\triangle t) &= \frac{1}{2\pi}\int_0^{2\pi}[-\phi(Q_1(\theta))\,\cos\theta + u(Q_1(\theta))\,\cos^2\theta + v(Q_1(\theta))\,\sin\theta\,\cos\theta]\,\mathrm{d}\theta \\
&\quad +\frac{1}{2}u(Q_2) + \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\triangle t}\cos\theta S(\underline{x}+c\tau\underline{n}(\theta),\tilde{t},\theta)\,\mathrm{d}\tau\,\mathrm{d}\theta \\
&\quad -\frac{1}{2}c\int_0^{\triangle t}\phi_x(\tilde{Q}_2)\,\mathrm{d}\tau, \qquad (4.11)
\end{aligned}$$

$$\begin{aligned}
v(\underline{x}, t+\triangle t) &= \frac{1}{2\pi}\int_0^{2\pi}[-\phi(Q_1(\theta))\,\sin\theta + u(Q_1(\theta))\,\cos\theta\,\sin\theta + v(Q_1(\theta))\,\sin^2\theta]\,\mathrm{d}\theta \\
&\quad +\frac{1}{2}v(Q_2) + \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\triangle t}\sin\theta S(\underline{x}+c\tau\underline{n}(\theta),\tilde{t},\theta)\,\mathrm{d}\tau\,\mathrm{d}\theta \\
&\quad -\frac{1}{2}c\int_0^{\triangle t}\phi_y(\tilde{Q}_2)\,\mathrm{d}\tau, \qquad (4.12)
\end{aligned}$$

where $S(\underline{x}+c\tau\underline{n}(\theta),\tilde{t},\theta) \equiv S(\tilde{t},\theta)$ is given by (2.4). We have obtained two exact representation formulae (2.6) – (2.8) and (4.10) – (4.12), which are equivalent. This can be easily verified by integrating the second and third equations of (2.1) from $Q_2 \equiv P'$ to P. As with Butler's scheme, the final integrals in (4.11) and (4.12) involving ϕ_x and ϕ_y need to be replaced by integrals over the cone mantle. In particular, we generate the evolution Galerkin scheme we have called EG3 from applying the rectangle rule to the time integrals in the above representation. Then, since $Q_2 \equiv P'$, the formulae to replace ϕ_x and ϕ_y at Q_2 are as in (4.4).

Approximate evolution operator for EG3

In the notation used above for EG1 and EG2, we obtain

$$\phi_P = \frac{1}{2\pi}\int_0^{2\pi}[\phi_Q - 2u_Q\cos\theta - 2v_Q\sin\theta]d\theta + O(\Delta t^2) \qquad (4.13)$$

$$u_P = \frac{1}{2}u_{P'} + \frac{1}{2\pi}\int_0^{2\pi}[-2\phi_Q\cos\theta + u_Q(3\cos^2\theta - 1) + 3v_Q\sin\theta\cos\theta]d\theta + \ O(\Delta t^2) \qquad (4.14)$$

$$v_P = \frac{1}{2}v_{P'} + \frac{1}{2\pi}\int_0^{2\pi}[-2\phi_Q\sin\theta + 3u_Q\sin\theta\cos\theta + v_Q(3\sin^2\theta - 1)]d\theta + \ O(\Delta t^2), \qquad (4.15)$$

where $Q = (x + c\Delta t \cos\theta, y + c\Delta t \sin\theta, t) \equiv Q_1$. These formulae are similar to those derived by Ostkamp in [20], [21] which she showed to be closely related to the method of transport developed independently by Fey [8] for the Euler equations.
The EG3 and Ostkamp schemes differ by the terms $u_Q(2\cos^2\theta - 1)$ and $2v_Q \sin\theta\cos\theta$ included in the integral (4.14), which are neglected in the Ostkamp scheme, along with similar terms in (4.15); clearly these terms are zero for constant u_Q and v_Q but are important when the base of the characteristic cone intersects neighbouring cells. [Note that in the preprint version of this paper and in [15] we did not distinguish between the Ostkamp and the EG3 schemes.]
A brief comparison of the three schemes EG1, EG2 and EG3 is now in order. They are all limited to first order accuracy by the projection onto piecewise constant approximations in the evolution Galerkin formulation (2.21). The first two are based on Butler's exact evolution operator given by (2.6) – (2.8); and the EG3 scheme on the exactly equivalent formulation of Ostkamp, given by (4.10) – (4.12). Then EG1 and EG3 lead to approximate evolution operators, which are only first order accurate in Δt, by use of a rectangle rule; while EG2 gives a second order accurate E_Δ by use of the trapezium rule in a similar way to Butler. However, in all the numerical tests described below in Section 6, EG3 consistently out performs not only EG1 but also EG2.
It is clear that the Ostkamp formulation of the evolution operator is significantly better than that of Butler. Moreover, it is not difficult to see why this is so. The wave system of equations (2.1) has the vorticity $\partial u/\partial y - \partial v/\partial x$ as a constant of the motion; and, from the characteristic form in this section, we see that the equivalent expression $u\sin\theta - v\cos\theta$ is carried up the cone axis. This fact is never exploited in the Butler formulation. But, although premultiplication of the characteristic variables by $\underline{\underline{R}}$ to get (4.10) – (4.12) obscures this property to some extent, it will be apparent from the numerical results in Section 6 that this is a very significant feature of the EG3 scheme. There still remains the possibility of exploiting the characteristic form, and in particular this property, more effectively than does the Ostkamp formulation.

4.4 A finite volume formulation

Despite the attractive features of the evolution operator given by (4.10) – (4.12), we are left with the difficulty of integrating the source terms to higher accuracy, and the experience of using the trapezium rule in deriving EG2 is not very encouraging. So in this section we briefly explore an alternative, finite volume, formulation. It is shown in [18], and the references therein, that for a scalar conservation law the use of the evolution operator in a finite volume formulation gives the same scheme as the direct formulation given above, because of the single family of characteristics. But we shall see that here we obtain different discrete schemes by this means.
If we integrate equation (4.8) over a mesh cell centred at point P, and over the time interval from $n\Delta t$ to $(n+1)\Delta t$, application of Gauss' formula gives

$$\underline{U}^{n+1} - \underline{U}^{n} + \Delta t \int_0^{\Delta t} \left[\underline{\underline{A}}_1 \delta_x \underline{U}^{n+\tau/\Delta t} + \underline{\underline{A}}_2 \delta_y \underline{U}^{n+\tau/\Delta t}\right] \mathrm{d}\tau = 0. \tag{4.16}$$

In this formula, $\underline{U}^n$ and $\underline{U}^{n+1}$ represent averages over a mesh cell at a given time level, while $\delta_x\underline{U}^{n+\tau/\Delta t}$ involves averages along the cell edges to the right and left and $\delta_y\underline{U}^{n+\tau/\Delta t}$ along edges to the top and bottom, in all cases at an intermediate time level $n\Delta t + \tau$. The approximate evolution operator is now used to evaluate these edge fluxes.

There are several advantages to this formulation. The most important is that the first order accurate approximation E_τ to the evolution operator $E(\tau)$ yields an overall second order update from $\underline{U}^n$ to $\underline{U}^{n+1}$. To obtain this second order approximation in the discrete scheme it is only necessary to carry out a recovery stage at each time level to generate a piecewise linear approximation $\tilde{\underline{U}}^n$, from the piecewise constant $\underline{U}^n$, to feed into the calculation of the fluxes. Moreover, it is at this stage that some solution dependence of the scheme can be introduced, for example to maintain monotonicity, positivity or TVD properties of the approximation - see [18] again.

We illustrate this procedure by returning to the special case of one-dimensional data discussed at the end of Section 2.3. For the CFL number $\nu \leq 1$, the edge values are independent of τ and for EG1 are given by

$$\phi^{n+*} = \mu_x \phi^n - \frac{2}{\pi}\delta_x u^n, \quad u^{n+*} = \mu_x u^n - \frac{2}{\pi}\delta_x \phi^n, \tag{4.17}$$

where we have extended the notation of (2.24) with $\mu_x f(x) := 1/2[f(x+h/2)+f(x-h/2)]$, $\delta_x f(x) := f(x+h/2) - f(x-h/2)$. Substitution into (4.16) then gives an update of the form (2.26), but with $\nu/2$ replaced by $2\nu/\pi$ for both the ϕ and the u update. So now there is slightly more damping than with the standard first order upwind scheme.

If, on the other hand, from the many possible recovery schemes we select one that approximates ϕ at an edge by a linear function that equals the average from the cells either side of the edge, i.e. $\mu_x\phi$, and with a slope given by $h^{-1}\delta_x\phi$, we obtain the edge values

$$\phi^{n+*} = \mu_x \phi^n - \frac{c\tau}{h}\delta_x u^n, \quad u^{n+*} = \mu_x u^n - \frac{c\tau}{h}\delta_x \phi^n. \tag{4.18}$$

Substituted into (4.16) this gives the second order accurate Lax-Wendroff scheme.

Thus we see that putting the EG1 approximate evolution operator into this finite volume formulation restores the common operator treatment of the ϕ and u variables, in this one-dimensional case; and it gives a scheme very close to that obtained from the exact evolution operator, cf. (4.1). We see that it will also give a second order accurate scheme when a recovery stage is added.

Further development of this approach, based on the evolution operator used for EG3, is deferred to a later paper because, not only is this the most convenient approach to obtaining second order accuracy, but it is also much to be preferred for imposing typical boundary conditions.

5 Stability and error analysis

We consider the general hyperbolic system (3.1) and denote by $\|\cdot\|$ the L^2-norm and by $\|\cdot\|_k$ the H^k-norm. The global error between the exact solution of the hyperbolic system considered and the approximate solution $\underline{U}^n$ is then defined as

$$\underline{e}^n := \underline{U}(t_n) - \underline{U}^n.$$

The error can be decomposed into a projection error $\underline{\eta}$ and an evolutionary error $\underline{\xi}$:

$$\underline{e}^n = (\underline{U}(t_n) - Q\underline{U}(t_n)) + (Q\underline{U}(t_n) - \underline{U}^n) \equiv \underline{\eta}^n + \underline{\xi}^n, \tag{5.1}$$

where $Q : L^2 \to S_h$ is a suitable projection onto S_h, possibily different from P_h.

In what follows we consider projections for which the following projection error estimates hold. Take $\underline{U}(t) \in (H^{r+1}(I\!\!R^d))^m$ then

$$\|\underline{\eta}\|_s \equiv \|\underline{U}(t) - Q\underline{U}(t)\|_s \leq C_1 h^{r+1-s} \|\underline{U}(t)\|_{r+1}, \tag{5.2}$$

for any $s \in [0, r]$. An example for such a projection will be given below. For a piecewise constant approximation (i.e. $r = 0$) the projection error in the L^2-norm is assumed to be $O(h)$; for a piecewise linear approximation (i.e. $r = 1$) it would be $O(h^2)$.
We derive an evolution equation for ξ by introducing the term $P_h E_\Delta Q\underline{U}$ as follows,

$$\underline{\xi}^n = (Q\underline{U}(t_n) - P_h E_\Delta Q\underline{U}(t_{n-1})) + (P_h E_\Delta Q\underline{U}(t_{n-1}) - P_h E_\Delta \underline{U}^{n-1}). \tag{5.3}$$

If the operator $P_h E_\Delta$ is strongly stable, i.e.

$$\|P_h E_\Delta\| \leq 1, \tag{5.4}$$

the last term is bounded by $\|\underline{\xi}^{n-1}\|$, so that the evolutionary error $\underline{\xi}^n$ is determined by the truncation error

$$T^n := \frac{1}{\Delta t}(Q\underline{U}(t_n) - P_h E_\Delta Q\underline{U}(t_{n-1})), \tag{5.5}$$

through the recurrence relation $\|\underline{\xi}^n\| \leq \|\underline{\xi}^{n-1}\| + \Delta t\|T^n\|$.
Generally, to obtain an order of accuracy p, we need both the projection error $\underline{\eta}^n$ and the truncation error T^n to be of this order. But we have a free choice of the projection Q to ensure that this holds. Therefore, it remains to establish the strong stability of $P_h E_\Delta$, to choose Q and to prove the order of the truncation error.

5.1 Strong stability of the EG1, EG2 and EG3 schemes

Our aim in this subsection will be to consider the stability of the evolution Galerkin schemes EG1, EG2 and EG3 for the wave equation system. The approximate evolution operators E_Δ are defined in (2.17), (2.18), (2.19); (4.5), (4.6), (4.7) and in (4.13), (4.14), (4.15) for the EG1, EG2 and EG3 schemes, respectively.
Using the L^2-projection P_h (cf. (2.22)) and the approximate evolution operators E_Δ the resulting finite difference formulae for each EG scheme can be put into the following form

$$\underline{U}_{kl}^{n+1} = \underline{U}_{kl}^n + \sum_{i=-1}^{1} \sum_{j=-1}^{1} \underline{\underline{C}}_{ij} \underline{U}_{k+i\, l+j}^n, \tag{5.6}$$

where the entries of the matrices $\underline{\underline{C}}_{ij}$ are taken appropriately according to the respective approximate evolution E_Δ. In the Appendix these are given in terms of stencil matrices $\underline{\underline{\alpha}}, \underline{\underline{\beta}}, \underline{\underline{\gamma}}$ where we have

$$\underline{\underline{C}}_{ij} = \begin{pmatrix} \alpha_{ij}^1 & \beta_{ij}^1 & \gamma_{ij}^1 \\ \alpha_{ij}^2 & \beta_{ij}^2 & \gamma_{ij}^2 \\ \alpha_{ij}^3 & \beta_{ij}^3 & \gamma_{ij}^3 \end{pmatrix}. \tag{5.7}$$

Due to the linearity of the problem it is possible (and most convenient) to use Fourier analysis to establish the stability of the EG schemes (5.6), i.e.to prove (5.4). It should be pointed out that this is nontrivial because of the approximations to the evolution operator that have been made: if it were possible to use the exact evolution operator

unconditional stability would follow immediately. On the other hand, as indicated in Section 2.3 the form of the scheme EG1 given there and of the EG2 and EG3 schemes given in the Appendix are those which are valid for the CFL number ν satisfying $\nu \leq 1$, so we need only to prove stability of the difference operators under this condition.
Even this is a difficult task; and to explore the form of the schemes and their stability for $\nu > 1$ would take us far beyond the scope of the present article. Fortunately, for these necessarily first order schemes (because they are obtained from a projection onto piecewise constants), the dominant damping term makes it easy to establish that there is a nonempty interval of stability; and this is what we shall do here. In a following paper [14] we shall discuss second order EG methods as well as standard second order methods such as Lax-Wendroff and Taylor-Galerkin methods. A more detailed analysis will be necessary and will be given there.

Lemma 5.1 *Each of the difference operators representing EG1, EG2 and EG3 is strongly stable in some positive interval* $0 \leq \nu < \nu_{max}$ *for the CFL number* ν.

Proof: For $0 \leq \nu \leq 1$, the schemes may be written as in (5.6) and the Appendix so that after Fourier transformation we have

$$\underline{\hat{U}}^{n+1} = [I - \nu(C_R + iC_I)]\, \underline{\hat{U}}^{n}, \tag{5.8}$$

where C_R, C_I are real symmetric matrices. Moreover, $C_R = D + \nu\tilde{C}_1$ and $C_I = \tilde{C}_2 + \nu\tilde{C}_3$, where D is a positive diagonal matrix, and $D, \tilde{C}_1, \tilde{C}_2$ and $\tilde{C}_3$ are independent of ν. Hence we can introduce a matrix K such that

$$\|\underline{\hat{U}}^{n+1}\|^2 = (\underline{\hat{U}}^{n})^* \left[I - 2\nu D + \nu^2 K\right] \underline{\hat{U}}^{n}. \tag{5.9}$$

Writing $s_x = \sin(\frac{1}{2}k_x h),\ s_y = \sin(\frac{1}{2}k_y h)$, it is easy to show that

$$\|D\| \geq d(s_x^2 + s_y^2), \qquad d > 0; \tag{5.10}$$

and, although K depends on ν, for $\nu \leq 1$ it is readily shown that

$$\|K\| \leq k(s_x^2 + s_y^2) \tag{5.11}$$

for some k, independently of k_x, k_y. It follows that $\|\underline{\hat{U}}^{n+1}\| \leq \|\underline{\hat{U}}^{n}\|$ for $\nu \leq \min(2d/k, 1)$.
□

5.2 Error estimates for the EG1, EG2 and EG3 schemes

In this part the truncation error estimates for the EG1, EG2 and EG3 schemes will be given. Together with the results above these will give error estimates for the schemes in cases where stability is established. It will be convenient to work with the projection $Q : (L^2(\mathbb{R}^d))^m \to S_h$ given by point evaluation, i.e.

$$Q\underline{U}(t_n) := \sum_{k,l\in\mathbb{Z}} \underline{U}(x_k, y_l, t_n)\chi_{kl} \qquad \underline{U} \in (L^2(\mathbb{R}^d))^m,$$

where $x_k = kh,\ y_l = lh$ is the midpoint of the square cell Ω_{kl}. We note that Q is a projection onto the space S_h of pieceweise constant functions such that (5.2) holds for $r = 0$, see Ciarlet [4]. This means that $\|\eta^n\| = O(h)$.

Lemma 5.2 (truncation error) *The truncation error T^n of the evolution Galerkin schemes EG1, EG2 and EG3 is of first order, i.e. $\|T^n\| = O(h)$ for $\Delta t/h = \lambda$ fixed.*

Proof: We evaluate and compare both terms of T^n in (5.5) assuming a smooth solution $\underline{U} \in \left(C^2(I\!R^d)\right)^m$. For $Q\underline{U}(t_n)$ we have by Taylor expansion

$$\underline{U}(x_k, y_l, t_n) = \underline{U}(x_k, y_l, t_{n-1}) + \Delta t \underline{U}_t(x_k, y_l, t_{n-1}) + O(\Delta t^2). \tag{5.12}$$

The wave equation system $\underline{U}_t = -\underline{\underline{A}}_1 \underline{U}_x - \underline{\underline{A}}_2 \underline{U}_y$ implies that for $(x, y) \in \Omega_{kl}$

$$\begin{aligned} Q\underline{U}(x, y, t_n) = \underline{U}(x_k, y_l, t_n) = \underline{U}(x_k, y_l, t_{n-1}) - \Delta t \underline{\underline{A}}_1 \underline{U}_x(x_k, y_l, t_{n-1}) \\ -\Delta t \underline{\underline{A}}_2 \underline{U}_y(x_k, y_l, t_{n-1}) + O(\Delta t^2). \end{aligned} \tag{5.13}$$

On the other hand we apply the numerical scheme $P_h E_\Delta Q$, in the form (5.6), to the exact solution $\underline{U}$ and use the Taylor expansion again. Then we have, using the definition of Q,

$$\begin{aligned} P_h E_\Delta Q\underline{U}(x, y, t_{n-1}) = \underline{U}(x_k, y_l, t_{n-1}) + \sum_{i,j=-1}^{1} \underline{\underline{C}}_{ij} \underline{U}(x_{k+i}, y_{l+j}, t_{n-1}) = \underline{U}(x_k, y_l, t_{n-1}) \\ + \sum_{i,j=-1}^{1} \underline{\underline{C}}_{ij} \left(\underline{U}(x_k, y_l, t_{n-1}) + ih\underline{U}_x(x_k, y_l, t_{n-1}) + jh\underline{U}_y(x_k, y_l, t_{n-1})\right) + O(h^2). \end{aligned} \tag{5.14}$$

Each of the schemes considered here satisfy **consistency conditions of first order**, i.e.

$$-\frac{\Delta t}{h}\underline{\underline{A}}_1 = \sum_{i,j=-1}^{1} i\underline{\underline{C}}_{ij}, \quad -\frac{\Delta t}{h}\underline{\underline{A}}_2 = \sum_{i,j=-1}^{1} j\underline{\underline{C}}_{ij}\,. \tag{5.15}$$

These can be verified by straightforward calculation using the stencil matrices $\underline{\underline{\alpha}}$, $\underline{\underline{\beta}}$, $\underline{\underline{\gamma}}$ from the Appendix. Thus,

$$P_h E_\Delta Q\underline{U}(x, y, t_{n-1}) = \underline{U}(x_k, y_l, t_{n-1}) - \Delta t \underline{\underline{A}}_1 \underline{U}_x(x_k, y_l, t_{n-1}) - \Delta t \underline{\underline{A}}_2 \underline{U}_y(x_k, y_l, t_{n-1}) + O(h^2). \tag{5.16}$$

Subtracting (5.16) from (5.13) and using the fact that we assume $\Delta t/h$ to be constant we have the estimate for the truncation error from (5.5)

$$\|T^n\| \le C_2 \frac{h^2 + \Delta t^2}{\Delta t} = O(h). \tag{5.17}$$

$\square$

Now we can give error estimates for the EG schemes.

Theorem 5.1 (error estimate) *The evolution Galerkin schemes EG1, EG2 and EG3 are of first order for $0 \le \nu \le \nu_{max}$, where $\nu_{max} \in (0, 1]$. Suppose we are interested in the approximation at time $t = T$ and $n = T/\Delta t$. Then there exists a constant $C > 0$, such that*

$$\|e^n\| \le Ch. \tag{5.18}$$

Proof: The estimate follows straightforwardly using (5.1), (5.2), (5.3), (5.4) and Lemmas 5.1 and 5.2

$$\|e^n\| \le \|\underline{\eta}^n\| + \|\underline{\xi}^n\| \le C_1 h + \Delta t \|T^n\| + \|P_h E_\Delta\| \|\underline{\xi}^{n-1}\| \qquad (5.19)$$
$$\le C_1 h + \Delta t \textstyle\sum_{k=1}^{n} \|T^k\|.$$

The fact that $n\Delta t = T$ gives

$$\|e^n\| \le C_1 h + T C_2 h \le Ch. \qquad (5.20)$$

□

6 Numerical results

We will present results of numerical experiments for the wave equation system with continuous and discontinous data. The EG schemes have been studied experimentally and compared with other well-known numerical methods.

Problem 1.
We consider the initial value problem for the wave equation with the initial values

$$\phi(\underline{x}, 0) = \frac{1}{c}(\sin 2\pi x + \sin 2\pi y), \;\; u(\underline{x}, 0) = 0 = v(\underline{x}, 0).$$

In this case the exact solution is known:

$$\phi(\underline{x}, t) = \frac{1}{c} \cos 2\pi ct (\sin 2\pi x + \sin 2\pi y), \qquad (6.1)$$

$$u(\underline{x}, t) = \frac{1}{c} \sin 2\pi ct \cos 2\pi x, \qquad (6.2)$$

$$v(\underline{x}, t) = \frac{1}{c} \sin 2\pi ct \cos 2\pi y. \qquad (6.3)$$

The general behaviour of the solution is plotted in Figure 2. The computational domain $[-1, 1] \times [-1, 1]$ was divided into 80×80 cells. The solution shown was obtained by the EG3 scheme at time $T = 0.2$.
In Tables 1,2,3 below the errors for the schemes EG1, EG2 and EG3 are given for meshes of 40×40, $80 \times 80, \ldots, 640 \times 640$ cells, together with the experimental order of convergence (EOC) computed from two meshes of sizes N_1 and N_2 as

$$EOC = \ln \frac{\|\underline{U}_{N_1}(T) - \underline{U}^n_{N_1}\|}{\|\underline{U}_{N_2}(T) - \underline{U}^n_{N_2}\|} / \ln\left(\frac{N_2}{N_1}\right).$$

In all cases the results are for a CFL-number ν of 0.55 and an end time $T = 0.4$, but experiments for several other values of ν and T have confirmed the comparative accuracies of the schemes; that is, EG3 consistently out-performs both EG1 and EG2.

N	$\|\phi(T) - \phi^n\|$	$\|u(T) - u^n\|$	$\|\underline{U}(T) - \underline{U}^n\|$	EOC
40	0.150773	0.060456	0.173327	
80	0.076730	0.031408	0.088659	0.967157
160	0.039043	0.016164	0.045242	0.970604
320	0.019692	0.00820	0.022853	0.985279

Table 1. EG1 scheme, $T = 0.4$, CFL $= 0.55$

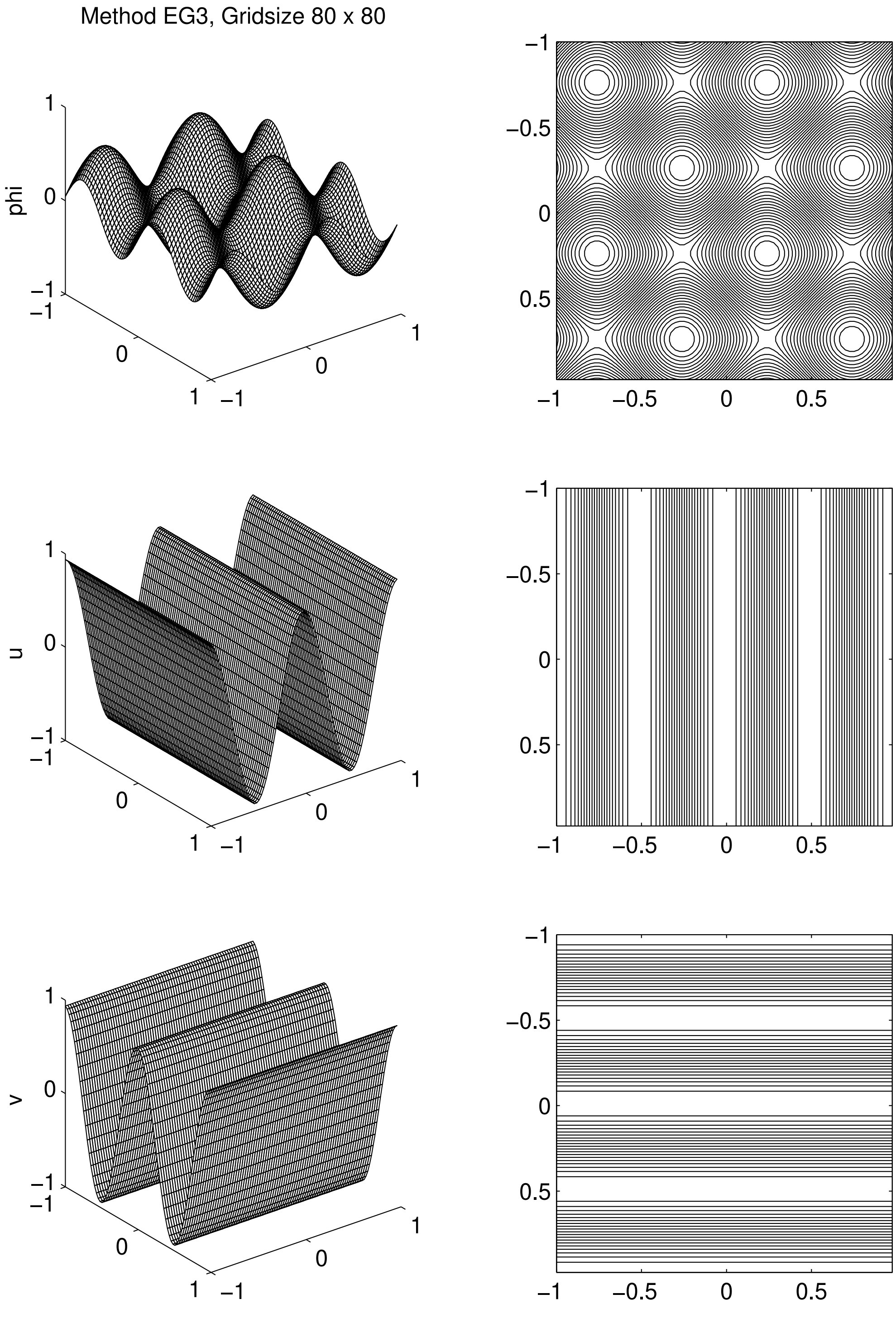

Figure 2: Components of the solution obtained by the EG3 scheme, their graphs and isolines

N	$\|\phi(T)-\phi^n\|$	$\|u(T)-u^n\|$	$\|\underline{U}(T)-\underline{U}^n\|$	EOC
40	0.164243	0.096602	0.213631	
80	0.087041	0.049397	0.111608	0.936681
160	0.045169	0.025042	0.057397	0.959393
320	0.023008	0.012599	0.029100	0.979956

Table 2. EG2 scheme, $T = 0.4$, CFL $= 0.55$

N	$\|\phi(T)-\phi^n\|$	$\|u(T)-u^n\|$	$\|\underline{U}(T)-\underline{U}^n\|$	EOC
40	0.0453708	0.004218	0.045761	
80	0.017574	0.004112	0.018511	1.305735
160	0.007876	0.002801	0.008813	1.070677
320	0.003709	0.001593	0.004340	1.022041

Table 3. EG3 scheme, $T = 0.4$, CFL $= 0.55$

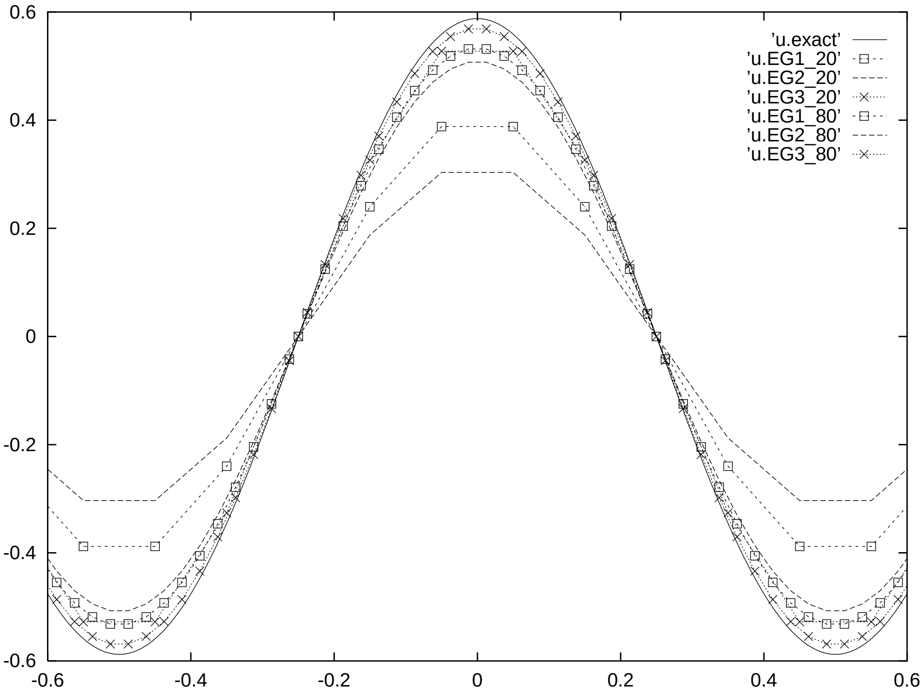

Figure 3: Velocity u(x,0,0.4) obtained by the EG schemes on different meshes

In Figure 3 the second component of the solution, the velocity $u(x, 0, T)$ restricted to the x - axis, for a 20×20 and an 80×80 mesh at the end time $T = 0.4$ is plotted. We see that EG1 and EG2 introduce much more damping of the wave than does EG3. This is, in fact, reflected in the numerical diffusion of the schemes, particularly the values of the coefficients a_{11}, $a_{22}(= a'_{33})$ and $a_{33}(= a'_{22})$ given in the Appendix. Thus the term $a_{11}+a_{22}$ has the relative values 9 : 11 : 6 in the three schemes, while EG1 is the only scheme which has $a_{33} \neq 0$, which boosts the sum $a_{11} + a_{22} + a_{33}$ to the relative value 12 in this case. For comparison, the first order upwind scheme discussed at the end of Section 2.3 would give a relative value of 3π for this sum, and the Lax-Wendroff scheme a value $3\pi\nu$.
To check that the experimental orders of convergence are reasonably representative we have also made other tests with linear and exponential initial conditions, see [15], as well

as for different values of ν. We are indeed able to introduce any piecewise linear initial data and give a comparison of our EG methods with a variety of other numerical methods for solving the wave equation system. The behaviour of the schemes as described above has been confirmed.

However, it should be noted that some schemes display exceptionally good accuracy for special values of ν, for example when the quantity $a_{11} + a_{22}$ discussed above is close to the Lax-Wendroff value $3\pi\nu$. In particular, for EG3 this occurs when $\nu = 2/\pi = 0.6366$, and for the Ostkamp scheme when $\nu = 5/3\pi = 0.5305$.

Problem 2.

As noted in Section 4.3, the wave equation system preserves the vorticity $\partial u/\partial y - \partial v/\partial x$. This is trivially zero in Problem 1, because u is independent of y and v of x; so now we take the following initial conditions, for which this is not true while the solution still has vanishing vorticity:

$$\phi(\underline{x}, 0) = c\exp(-10x^2 - 10y^2), \;\; u(\underline{x}, 0) = 0 = v(\underline{x}, 0). \tag{6.4}$$

We compute the discrete vorticity DV given by the formula

$$DV_{kl} := \mu_x \delta_y u_{k+1/2,l+1/2} - \mu_y \delta_x v_{k+1/2,l+1/2}, \qquad \text{for each } k, l \in \mathbb{Z}. \tag{6.5}$$

where by $u_{k\pm 1/2\, l\pm 1/2} := u((k \pm 1/2)h, (l \pm 1/2)h)$ we denote values at the corner points of the square mesh Ω_{kl}. In Table 4 we show reference values for DV_{kl}, namely, the average value (vor-aver), the minimum (vor-min) and the maximum (vor-max). These are all after a time $T = 0.2$, with $N = 40$ and $N = 200$ mesh cells in each direction, and $\nu = 0.55$.

$N = 40$	EG1	EG2	EG3
vor-aver	$3.9478950 \cdot 10^{-4}$	$2.6959134 \cdot 10^{-4}$	$2.4984926 \cdot 10^{-4}$
vor-min	$-1.69011228 \cdot 10^{-3}$	$-1.10278088 \cdot 10^{-3}$	$-1.07928549 \cdot 10^{-3}$
vor-max	$1.69011228 \cdot 10^{-3}$	$1.10278088 \cdot 10^{-3}$	$1.07928549 \cdot 10^{-3}$

$N = 200$	EG1	EG2	EG3
vor-aver	$1.421346 \cdot 10^{-5}$	$9.53898 \cdot 10^{-6}$	$7.51300 \cdot 10^{-6}$
vor-min	$-6.616513 \cdot 10^{-5}$	$-4.397805 \cdot 10^{-5}$	$-3.497483 \cdot 10^{-5}$
vor-max	$6.616513 \cdot 10^{-5}$	$4.397805 \cdot 10^{-5}$	$3.497483 \cdot 10^{-5}$

Table 4. Vorticity preservation, CFL= 0.55

The table shows clearly that EG3 is better than EG1 and EG2 in preserving the initial zero vorticity. In [19], however, it is shown that there are nine-point difference schemes that exactly preserve the discrete vorticity given by (6.5). The "rotated-Richtmyer" variant of the Lax-Wendroff scheme, which is given in the Appendix, is the most familiar of these; but there is a first order scheme which is directly comparable to EG1, EG2 and EG3 which also preserves this quantity exactly. It is given in the Appendix where it is denoted by VOR.

In Table 5 we therefore compare the accuracy of this scheme with our three evolution Galerkin schemes. We make this comparison on the practical mesh $N = 40$ at the time $T = 0.2$ and again with $\nu = 0.55$; the difference from a fine mesh Lax-Wendroff calculation is given, and a 40×40 Lax-Wendroff calculation, denoted by LW*, is also included. We see that VOR has a similar accuracy to EG1 but that EG3 is significantly better; it is indeed quite comparable with the Lax-Wendroff scheme on this mesh.

	$\|$ EG1 - LW $\|$	$\|$ EG2 - LW $\|$	$\|$ EG3 - LW $\|$	$\|$ LW* - LW $\|$	$\|$ VOR - LW $\|$
$\underline{U}$	0.029777	0.037713	0.023824	0.024059	0.032352
phi	0.014338	0.017933	0.011517	0.012350	0.013936
u	0.018454	0.023460	0.014747	0.014601	0.020646

Table 6. Accuracy of the EG schemes and of the first order vorticity preserving scheme

Problem 3.
The next example contains a discontinuity in the initial data. Although the theory is not applicable to discontinuous data, we consider this problem since many physically relevant problems in gas dynamics have discontinous solutions; discontinuities in linear problems correspond to contact discontinuities in gas dynamics. We consider the data

$$\phi(\underline{x}, 0) = 0,$$

$$v(\underline{x}, 0) = u(\underline{x}, 0) = \frac{1}{\sqrt{2}} \begin{cases} 1, & |y| < |x|, \\ -1, & \text{elsewhere.} \end{cases}$$

In Figures 4 and 5 the isolines of the computed approximate solutions for the evolution Galerkin schemes EG1, EG2, EG3 and the Butler sheme are shown. The computational domain $[-1, 1] \times [-1, 1]$ was divided into 400×400 cells, the final time set to be $T = 0.4$. For the first component ϕ, two discontinuities propagate in the positive and negative direction of the main diagonal while for the velocities u and v an additional discontinuity occurs along this diagonal.
It can again be observed that the EG2 scheme has considerable numerical dissipation and discontinuities are smeared out. The second order scheme of Butler resolves shocks much better, but it produces oscillations. The EG3 scheme is clearly the best of the four schemes shown. Its good features have also been demonstrated in [15], where computations with the Euler equations have been compared with those produced by some directional splitting finite volume schemes; in particular, they demonstrate the good preservation of circular symmetry that we have seen here in the ϕ plot of Figure 5 as well as in Problem 2.

7 Conclusions

In this paper we have derived three evolution Galerkin schemes EG1, EG2 and EG3 for systems of hyperbolic equations in two space dimensions. They approximate the evolution of the solution through one time step by means of the bicharacteristics and then project this onto a finite element space of piecewise constant functions.
The first two schemes, EG1 and EG2, are based on a straightforward use of the bicharacteristics following the approach of Butler [3]. But the much more effective EG3 scheme is based on the quasidiagonalization of a general hyperbolic system introduced by Ostkamp [20]. All the schemes differ from the finite difference schemes of Butler [3] and Reddy *et al.* [24] by replacing the overlapping biquadratic interpolation by the L^2 projection onto a consistent piecewise constant approximation. Thus they retain the upwind features of the basic first order upwind approximation of the advection equation, to which they are related.
In this first paper introducing these schemes we have concentrated on their application to the first order wave equation system. A theoretical analysis shows that they are stable

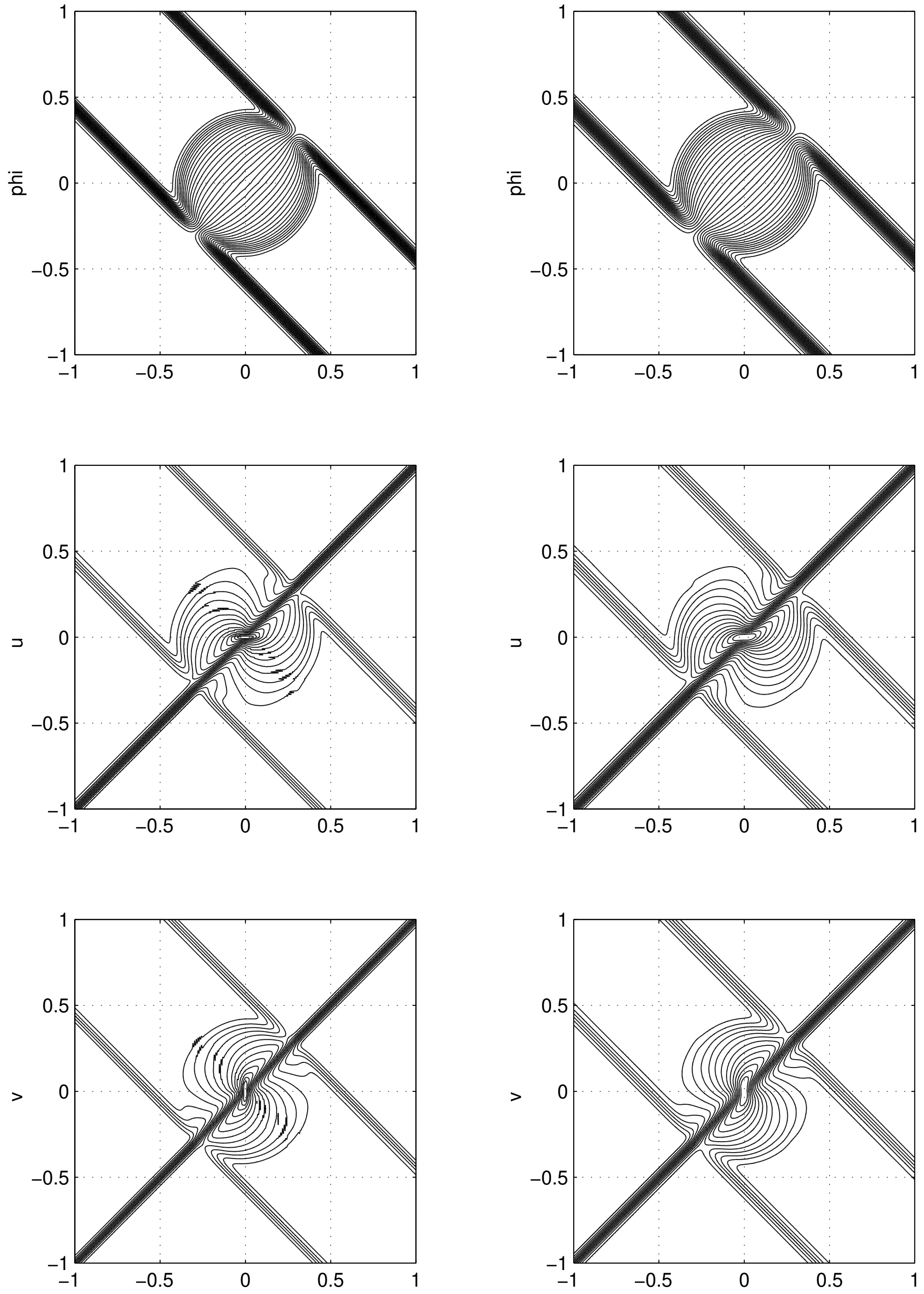

Figure 4: Isolines of the solution obtained by the EG1 and EG2 schemes

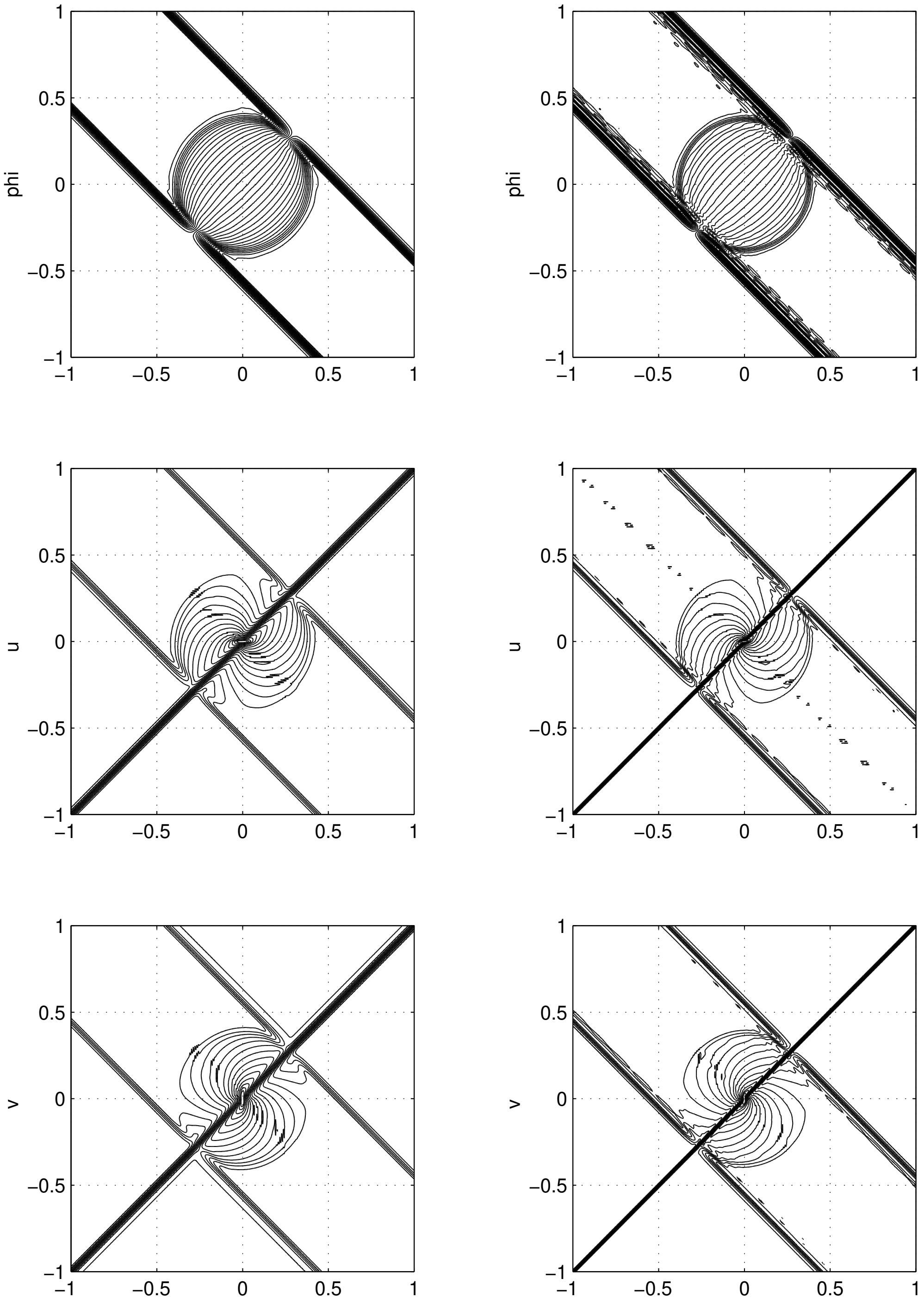

Figure 5: Isolines of the solution obtained by the EG3 and Butler schemes

for a range of CFL numbers $0 \leq \nu \leq \nu_{\max}$ and are first order accurate, both results being sharper than those obtained by Ostkamp [20], [21]. Numerical experiments demonstrate the superiority of EG3 over EG1 and EG2, good maintenance of solution properties such as circular symmetry independent of the mesh orientation, vorticity preservation, nonoscillatory approximation of discontinuous solutions and an accuracy that matches that of second order schemes on practical meshes.
In subsequent papers we intend to extend the schemes to second order accuracy, by use of a technique which is outlined here; we will apply more general boundary conditions and extend the stability analysis.

8 Appendix

We write out the finite difference formulation of all the numerical schemes we have referred to. Using the standard finite difference notation (cf. (2.24)) each scheme can be written down in the following form

$$\begin{aligned}
\phi^{n+1} &= \left[1 + a_{11}(\delta_x^2 + \delta_y^2) + b_{11}\delta_x^2\delta_y^2\right]\phi^n - \nu(1 + a_{12}\delta_y^2)\Delta_{0x}u^n - \nu(1 + a_{13}\delta_x^2)\Delta_{0y}v^n \\
u^{n+1} &= \left[1 + (a_{22}\delta_x^2 + a'_{22}\delta_y^2) + b_{22}\delta_x^2\delta_y^2\right]u^n - \nu(1 + a_{21}\delta_y^2)\Delta_{0x}\phi^n + \nu^2 a_{23}\Delta_{0x}\Delta_{0y}v^n \\
v^{n+1} &= \left[1 + (a_{33}\delta_x^2 + a'_{33}\delta_y^2) + b_{33}\delta_x^2\delta_y^2\right]v^n - \nu(1 + a_{31}\delta_x^2)\Delta_{0y}\phi^n + \nu^2 a_{32}\Delta_{0x}\Delta_{0y}u^n.
\end{aligned}$$

Below the coefficients of the schemes are given. For simplicity we use the notation $\nu = \frac{c\Delta t}{h}$.

- **EG1 scheme**

$$\begin{aligned}
&a_{11} = \nu/\pi, \quad b_{11} = \nu^2/4\pi, \quad a_{12} = 2\nu/3\pi, \quad a_{13} = 2\nu/3\pi, \\
&a_{21} = 2\nu/3\pi, \quad a_{22} = 2\nu/\pi, \quad a'_{22} = 0, \quad b_{22} = \nu^2/4\pi, \quad a_{23} = 3/4, \\
&a_{31} = 2\nu/3\pi, \quad a_{32} = 3/4, \quad a_{33} = 0, \quad a'_{33} = 2\nu/\pi, \quad b_{33} = \nu^2/4\pi
\end{aligned}$$

- **EG2 scheme**

$$\begin{aligned}
&a_{11} = 2\nu/\pi, \quad b_{11} = \nu^2/2\pi, \quad a_{12} = 2\nu/3\pi, \quad a_{13} = 2\nu/3\pi, \\
&a_{21} = 2\nu/3\pi, \quad a_{22} = 5\nu/3\pi, \quad a'_{22} = \nu/3\pi, \quad b_{22} = \nu^2/4\pi, \quad a_{23} = 1/2, \\
&a_{31} = 2\nu/3\pi, \quad a_{32} = 1/2, \quad a_{33} = \nu/3\pi, \quad a'_{33} = 5\nu/3\pi, \quad b_{33} = \nu^2/4\pi
\end{aligned}$$

- **EG3 scheme**

$$\begin{aligned}
&a_{11} = \nu/\pi, \quad b_{11} = \nu^2/4\pi, \quad a_{12} = 2\nu/3\pi, \quad a_{13} = 2\nu/3\pi, \\
&a_{21} = 2\nu/3\pi, \quad a_{22} = \nu/\pi, \quad a'_{22} = 0, \quad b_{22} = \nu^2/8\pi, \quad a_{23} = 3/8, \\
&a_{31} = 2\nu/3\pi, \quad a_{32} = 3/8, \quad a_{33} = 0, \quad a'_{33} = \nu/\pi, \quad b_{33} = \nu^2/8\pi
\end{aligned}$$

- **Ostkamp scheme**

$$\begin{aligned}
&a_{11} = \nu/\pi, \quad b_{11} = \nu^2/4\pi, \quad a_{12} = 2\nu/3\pi, \quad a_{13} = 2\nu/3\pi, \\
&a_{21} = 2\nu/3\pi, \quad a_{22} = 2\nu/3\pi, \quad a'_{22} = \nu/3\pi, \quad b_{22} = \nu^2/8\pi, \quad a_{23} = 1/8, \\
&a_{31} = 2\nu/3\pi, \quad a_{32} = 1/8, \quad a_{33} = \nu/3\pi, \quad a'_{33} = 2\nu/3\pi, \quad b_{33} = \nu^2/8\pi
\end{aligned}$$

- **VOR scheme**

$$
\begin{aligned}
&a_{11}=\nu/2, \quad b_{11}=\nu/4, \quad a_{12}=1/4, \quad a_{13}=1/4,\\
&a_{21}=1/4, \quad a_{22}=\nu/2, \quad a'_{22}=0, \quad b_{22}=\nu/8, \quad a_{23}=1/(2\nu),\\
&a_{31}=1/4, \quad a_{32}=1/(2\nu), \quad a_{33}=0, \quad a'_{33}=\nu/2, \quad b_{33}=\nu/8
\end{aligned}
$$

- **Lax-Wendroff scheme**

$$
\begin{aligned}
&a_{11}=\nu^2/2, \quad b_{11}=\nu^2/4, \quad a_{12}=1/4, \quad a_{13}=1/4,\\
&a_{21}=1/4, \quad a_{22}=\nu^2/2, \quad a'_{22}=0, \quad b_{22}=\nu^2/8, \quad a_{23}=1/2,\\
&a_{31}=1/4, \quad a_{32}=1/2, \quad a_{33}=0, \quad a'_{33}=\nu^2/2, \quad b_{33}=\nu^2/8
\end{aligned}
$$

- **Butler scheme**

$$
\begin{aligned}
&a_{11}=\nu^2/2, \quad b_{11}=0, \quad a_{12}=\nu^2/4, \quad a_{13}=\nu^2/4,\\
&a_{21}=0, \quad a_{22}=\nu^2/2, \quad a'_{22}=0, \quad b_{22}=0, \quad a_{23}=1/2,\\
&a_{31}=0, \quad a_{32}=1/2, \quad a_{33}=0, \quad a'_{33}=\nu^2/2, \quad b_{33}=0
\end{aligned}
$$

- **Prasad scheme**

$$
\begin{aligned}
&a_{11}=\nu^2/2, \quad b_{11}=\nu^4/12, \quad a_{12}=\nu^2/4, \quad a_{13}=\nu^2/4,\\
&a_{21}=\nu^2/6, \quad a_{22}=\nu^2/2, \quad a'_{22}=0, \quad b_{22}=\nu^4/24, \quad a_{23}=1/2,\\
&a_{31}=\nu^2/6, \quad a_{32}=1/2, \quad a_{33}=0, \quad a'_{33}=\nu^2/2, \quad b_{33}=\nu^4/24
\end{aligned}
$$

In Section 5.1 we have used a finite difference formulation in the following form:

$$
\begin{aligned}
\phi_{kl}^{n+1} &= \phi_{kl}^{n} + \sum_{i,j=-1}^{1}\left(\alpha_{ij}^{1}\phi_{k+i\,l+j}^{n} + \beta_{ij}^{1}u_{k+i\,l+j}^{n} + \gamma_{ij}^{1}v_{k+i\,l+j}^{n}\right)\\
u_{kl}^{n+1} &= u_{kl}^{n} + \sum_{i,j=-1}^{1}\left(\alpha_{ij}^{2}\phi_{k+i\,l+j}^{n} + \beta_{ij}^{2}u_{k+i\,l+j}^{n} + \gamma_{ij}^{2}v_{k+i\,l+j}^{n}\right)\\
v_{kl}^{n+1} &= v_{kl}^{n} + \sum_{i,j=-1}^{1}\left(\alpha_{ij}^{3}\phi_{k+i\,l+j}^{n} + \beta_{ij}^{3}u_{k+i\,l+j}^{n} + \gamma_{ij}^{3}v_{k+i\,l+j}^{n}\right).
\end{aligned}
$$

Below the stencil matrices $\underline{\underline{\alpha}}, \underline{\underline{\beta}}, \underline{\underline{\gamma}}$ for the EG3 scheme are given:

EG3 scheme

$$\underline{\underline{\alpha}}^1 := \left\{ \begin{array}{ccc} \frac{\nu^2}{4\pi} & \frac{\nu}{\pi} - \frac{\nu^2}{2\pi} & \frac{\nu^2}{4\pi} \\ \frac{\nu}{\pi} - \frac{\nu^2}{2\pi} & -\frac{4\nu}{\pi} + \frac{\nu^2}{\pi} & \frac{\nu}{\pi} - \frac{\nu^2}{2\pi} \\ \frac{\nu^2}{4\pi} & \frac{\nu}{\pi} - \frac{\nu^2}{2\pi} & \frac{\nu^2}{4\pi} \end{array} \right\}, \quad \underline{\underline{\beta}}^1 := \left\{ \begin{array}{ccc} \frac{\nu^2}{3\pi} & 0 & \frac{-\nu^2}{3\pi} \\ \frac{\nu}{2} - \frac{2\nu^2}{3\pi} & 0 & -\frac{\nu}{2} + \frac{2\nu^2}{3\pi} \\ \frac{\nu^2}{3\pi} & 0 & -\frac{\nu^2}{3\pi} \end{array} \right\},$$

$$\underline{\underline{\gamma}}^1 := \left\{ \begin{array}{ccc} -\frac{\nu^2}{3\pi} & -\frac{\nu}{2} + \frac{2\nu^2}{3\pi} & -\frac{\nu^2}{3\pi} \\ 0 & 0 & 0 \\ \frac{\nu^2}{3\pi} & \frac{\nu}{2} - \frac{2\nu^2}{3\pi} & \frac{\nu^2}{3\pi} \end{array} \right\}, \quad \underline{\underline{\alpha}}^2 := \left\{ \begin{array}{ccc} \frac{\nu^2}{3\pi} & 0 & -\frac{\nu^2}{3\pi} \\ \frac{\nu}{2} - \frac{2\nu^2}{3\pi} & 0 & -\frac{\nu}{2} + \frac{2\nu^2}{3\pi} \\ \frac{\nu^2}{3\pi} & 0 & \frac{-\nu^2}{3\pi} \end{array} \right\},$$

$$\underline{\underline{\beta}}^2 := \left\{ \begin{array}{ccc} \frac{\nu^2}{8\pi} & -\frac{\nu^2}{4\pi} & \frac{\nu^2}{8\pi} \\ \frac{\nu}{\pi} - \frac{\nu^2}{4\pi} & -\frac{2\nu}{\pi} + \frac{\nu^2}{2\pi} & \frac{\nu}{\pi} - \frac{\nu^2}{4\pi} \\ \frac{\nu^2}{8\pi} & -\frac{\nu^2}{4\pi} & \frac{\nu^2}{8\pi} \end{array} \right\}, \quad \underline{\underline{\gamma}}^2 := \left\{ \begin{array}{ccc} -\frac{3\nu^2}{32} & 0 & \frac{3\nu^2}{32} \\ 0 & 0 & 0 \\ \frac{3\nu^2}{32} & 0 & -\frac{\nu^2}{32} \end{array} \right\},$$

$$\underline{\underline{\alpha}}^3 := \left\{ \begin{array}{ccc} -\frac{\nu^2}{3\pi} & -\frac{\nu}{2} + \frac{2\nu^2}{3\pi} & -\frac{\nu^2}{3\pi} \\ 0 & 0 & 0 \\ \frac{\nu^2}{3\pi} & +\frac{\nu}{2} - \frac{2\nu^2}{3\pi} & \frac{\nu^2}{3\pi} \end{array} \right\}, \quad \underline{\underline{\beta}}^3 := \left\{ \begin{array}{ccc} -\frac{3\nu^2}{32} & 0 & \frac{3\nu^2}{32} \\ 0 & 0 & 0 \\ \frac{3\nu^2}{32} & 0 & -\frac{3\nu^2}{32} \end{array} \right\},$$

$$\underline{\underline{\gamma}}^3 := \left\{ \begin{array}{ccc} \frac{\nu^2}{8\pi} & \frac{\nu}{\pi} - \frac{\nu^2}{4\pi} & \frac{\nu^2}{8\pi} \\ -\frac{\nu^2}{4\pi} & -\frac{2\nu}{\pi} + \frac{\nu^2}{2\pi} & -\frac{\nu^2}{4\pi} \\ \frac{\nu^2}{8\pi} & \frac{\nu}{\pi} - \frac{\nu^2}{4\pi} & \frac{\nu^2}{8\pi} \end{array} \right\}.$$

Acknowledgements.
This research was supported under the DFG Grant No. Wa 633/6-1 of Deutsche Forschungsgemeinschaft and partially by the Grant No. 201/97/0153 of the Czech Grant Agency. We would like to acknowledge several clarifying discussions with Thomas Sonar and access to the program of Stella Ostkamp. The first author would like to thank Alexander Ženíšek for his continual encouragement and support of her research.
We would like to thank the referees for a large number of helpful comments which enabled us to improve the paper substantially.

References

[1] J.P. Benqué, G. Labadie, and J. Ronat. A new finite element method for the Navier-Stokes equations coupled with a temperature equation. In T. Kawai, editor, *Proceedings of the Fourth International Symposium on Finite Element Methods in Flow Problems*, pages 295–301. North-Holland, 1982.

[2] A. Brandt and S. Ta'asan. Multigrid solutions to quasi-elliptic schemes. In Earll. M. Murman and Saul Abarbanel, editors, *Progress and Supercomputing in Computational Fluid Dynamics*, Proceedings of the U.S.-Israel Workshop 1984, Birkhauser 1985.

[3] D.S. Butler. The numerical solution of hyperbolic systems of partial differential equations in three independent variables. *Proc. Roy. Soc.*, 255A:233–252, 1960.

[4] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems.* North–Holland, Amsterdam, 1979.

[5] P.N. Childs and K.W. Morton. Characteristic Galerkin methods for scalar conservation laws in one dimension. *SIAM J. Numer. Anal.*, 27:553–594, 1990.

[6] M.C. Cline and J.D. Hoffman. Comparison of characteristic schemes for three-dimensional, steady, isentropic flow. *AIAA J.*, 10(11):1452–1458, 1972.

[7] J. Douglas and T.F. Russell. Numerical methods for convection–dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures. *SIAM J. Numer. Anal.*, 19:871–885, 1982.

[8] M. Fey. *Ein echt mehrdimensionales Verfahren zur Lösung der Eulergleichungen*, 1993. Dissertation, ETH Zürich.

[9] M. Fey and R. Jeltsch. A simple multidimensional Euler scheme, *Proceedings of the First European Computational Fluid Dynamics Conference, ECCOMAS'92*, Vol.I, Ch.Hirsch et.al. (Editors), Elsevier Science Publishers, Amsterdam, 1992.

[10] T.N. Krishnamurti. Numerical integration of primitive equations by a quasi-Lagrangian advective scheme. *J. Appl. Meteorology*, 1:508–521, 1962.

[11] P. Lesaint. Numerical solution of the equation of continuity. In J.J.H. Miller, editor, *Topics in Numerical Analysis III*, pages 199–222. Academic Press, 1977.

[12] P. Lin, K.W. Morton and E. Süli. Euler characteristic Galerkin scheme with recovery. M^2AN, 27(7):863–894, 1993.

[13] P. Lin, K.W. Morton and E. Süli. Characteristic Galerkin schemes for scalar conservation laws in two and three space dimensions. *SINUM*, 34(2):779–796, 1997.

[14] M. Lukáčová - Medviďová, K.W. Morton, and G. Warnecke. The second order evolution Galerkin schemes for hyperbolic systems. *In preparation.*

[15] M. Lukáčová–Medviďová, K.W. Morton, and G. Warnecke, On the evolution Galerkin method for solving multidimensional hyperbolic systems, appears in *Proceedings of the Second European Conference on Numerical Mathematics and Advanced Applications* ENUMATH'97, Heidelberg,1998.

[16] K.W. Morton. Approximation of multidimensional hyperbolic partial differential equations. In I.S. Duff and G.A. Watson, editors, *Procedings of the State of the Art in Numerical Analysis Conference*, OUP, 473-502, 1997.

[17] K.W. Morton. *Numerical Solution of Convection-Diffusion Problems*, volume 12 of *Applied Mathematics and Mathematical Computation.* Chapman & Hall, London, 1996.

[18] K.W. Morton. On the analysis of finite volume methods for evolutionary problems, to appear in *SIAM J. Numer. Anal.*

[19] K.W. Morton and P.L. Roe. Vorticity-preserving Lax-Wendroff type schemes for the system wave equation. *In preparation.*

[20] S. Ostkamp. *Multidimensional characterisitic Galerkin schemes and evolution operators for hyperbolic systems.* PhD thesis, Universität Hannover, 1995.

[21] S. Ostkamp. Multidimensional characterisitic Galerkin schemes and evolution operators for hyperbolic systems, *Math. Meth. Appl. Sci.*, 20:1111-1125, 1997.

[22] O. Pironneau. On the transport-diffusion algorithm and its application to the Navier-Stokes equations. *Numer. Math.*, 38:309–332, 1982.

[23] P. Prasad and R. Ravindran. Canonical form of a quasilinear hyperbolic system of first order equations. *J. Math. Phys. Sci.*, 18(4):361–364, 1984.

[24] A.S. Reddy, V.G. Tikekar, and P. Prasad. Numerical solution of hyperbolic equations by method of bicharacteristics. *Journal of Mathematical and Physical Sciences*, 16(6):575–603, 1982.

[25] A. Staniforth and J. Côté. Semi-Lagrangian integration schemes and their application to environmental flows. *Monthly Weather Rev.*, 119(9):2206–2223, 1991.

[26] E. Süli. Convergence and nonlinear stability of the Lagrange-Galerkin method for the Navier-Stokes equations. *Numer. Math.*, 53:459–483, 1988.

[27] M. Tanguay, A. Simard, and A. Staniforth. A three-dimensional semi-Lagrangian scheme for the Canadian regional finite-element forecast model. *Monthly Weather Rev.*, 117:1861–1871, 1989.