Is Discontinuous Reconstruction Really a Good Idea?

Philip Roe

Abstract It has been almost automatically assumed for a quarter century that the numerical solution of hyperbolic conservation laws is best accomplished by making a reconstruction of the initial data that is only piecewise continuous. The effect of the discontinuities is taken into account by means of Riemann solvers. This strategy has enjoyed great practical success but introduces only one-dimensional physics as a guide to the discretization of multidimensional problems. This article points out some of the resulting defects and proposes an alternative viewpoint. The chief novelty of the new “Active Flux” method, apart from the elimination of discontinuities, is the division into advective and acoustic disturbances, with acoustics being handled by exploiting classical solutions to the scalar wave equation. Results for a standard test problem for the compressible Euler equations indicate that an order of magnitude increase in cost effectiveness may be possible by following this approach, compared to the traditional one.

Keywords Hyperbolic conservation laws · Numerical advection · Compressible flow · Wave equations

1 Some History, By Way of Introduction

In 1959, there appeared in the Russian mathematics journal Matematicheskii Sbornik, (Mathematical Miscellany) a paper that has exerted, directly or indirectly, a profound influence on almost all of the methods subsequently used to find numerical solutions of hyperbolic conservation laws. These include the so-called Godunov-type finite-volume methods and
their successors, such as discontinuous Galerkin and WENO schemes. In his paper [13] Sergei Godunov showed that no numerical method for the linear scalar advection equation \( \partial_t u + a \partial_x u = 0 \) can be both monotone and better than first-order accurate. He also showed that the linear monotone advection scheme with smallest truncation error was the simplest upwind scheme, and offered a plausible generalization to one-dimensional systems which is now well-known as Godunov’s method. This consists of projecting the initial data into the space of cellwise constant functions and then solving that modified problem exactly for a small interval of time. Then the new solution is again represented by piecewise constant functions and the cycle is repeated. At first, this proposal was not taken very seriously by practical scientists and engineers, who tended to see it as artificial, complicated and expensive. Understanding and acceptance were eased by the explanation that it did not attempt to find a discrete evolution operator, but instead worked by putting the data in a form to which the exact operator could be applied, at least for short times.

Godunov [14] has written an interesting account of the circumstances under which this work was done. He and his colleagues seem to have attached great importance to the interpretation given above, and they based their approach to two-dimensional problems on taking it as literally as possible, even though this forced them to confront a considerable difficulty. To quote…

The first question we faced in our attempts to generalize our method for two dimensions was how to find a solution of the two-dimensional Riemann problem with arbitrary initial conditions.… In order for a two-dimensional method to be absolutely similar to the one-dimensional one we needed an analytical solution of the gas-dynamic equations with four initial discontinuities coming together in a single point. Naturally, we did not have these solutions at that time and they are still unknown (for general initial conditions). At this point, a rogueish suggestion was made which was to use only the solutions of a classical Riemann problem involving only planar waves. Thus, the interaction of the four cells with a common vertex was neglected altogether. This removed a nice physical interpretation underlying the construction of the one-dimensional scheme. Quite naturally, there were many arguments during the discussion of this hardly justifiable suggestion. L. V. Brushlinkii… carried out the analytical solution for an acoustic wave propagating in stationary media, which took into account the interaction of the cells sharing a common vertex. His solution was implemented in a scheme completely analogous to the one-dimensional one. To our surprise and pleasure there were no significant differences. Afterwards, only the simpler model was used.

The simpler version that is in general use today, based solely on one-dimensional Riemann problems at cell edges, seems to have been a desperate measure resorted to only after the apparently more physically complete model that included corner interactions had proved both difficult and fruitless. Today, the irrelevance of the corner interactions is less surprising, because we are well aware what the solutions to “corner Riemann problems” look like, and it is hard to imagine that their addition, which would resemble Fig. 1, might improve anything.

The story is a little more complicated because several people, for example [6,21] have found that including corner interactions DOES improve the results from an upwind code for two-dimensional linear advection. But apparently that effect does not carry over to the Euler equations. The seeming paradox is resolved by realizing that advection, in any number of dimensions, is a scalar problem with a one-dimensional domain of dependence that is easily identified and taken into consideration. Acoustic waves, on the other hand, behave one-
dimensionally only in one dimension. In $n$ dimensions they have an $n$-dimensional domain of dependence that does not have a simple mapping onto the grid, and cannot be represented by any finite number of one-dimensional interactions. It follows that advection is not a reliable guide in more than one dimension if we actually wish to study acoustics, or systems such as the Euler equations that contain acoustic behavior.

In many recent attempts to formulate a computational framework for the Euler equations, it has been regarded as almost axiomatic that the discrete representation should be discontinuous.\footnote{A significant exception is the class of methods deriving from the SUPG concept put forward by Hughes and coworkers \cite{16,17}. These are not considered here as being outside our main interest.} Any consistent prediction of the evolution then involves considering one-dimensional waves propagating normal to the discontinuity, and therefore introduces Riemann solvers. Almost inevitably, two discontinuities will intersect, and their interaction will either be incorporated at considerable expense with little effect, or will in most cases be discarded with the admission that the physical model is incomplete.

\section{What Else Could There Be?}

“Multidimensional upwind” methods too numerous to cite have been published over the years, including attempts of my own \cite{32,33,37}. I give an incomplete summary of some of this work in \cite{35}. Most of this is still within the framework of discontinuous representations, however, and therefore still relies on one-dimensional flow models. In one dimension, the mathematical description of advective and acoustic phenomena is only distinguished by nonlinearity, so that other very real and important differences between them cannot properly be detected.

An attempt to avoid these limitations is to be found in the class of Residual Distribution, or Fluctuation-splitting schemes, introduced in \cite{34}. In the earliest versions, designed merely to solve advection problems, the solution was represented in two dimensions by unique values at the vertices of a triangular mesh, and the flux integral $\phi^T$ calculated around the boundary.
of each element $T$. This is divided into three, or fewer, unequal parts. These are used to update the vertex values of the element. Each vertex may receive updates from several of the elements around it. The rules for making the divisions and updates take into account both the flow direction and the element geometry. The flow is assumed to be one-dimensional, but not necessarily in a direction related to the element. This produced results with much crisper discontinuities than those from grid-aligned upwinding [8]. The generalization of these methods, first to systems of equations, then to unsteady flows, and eventually to viscous flows [1] has however followed a path that I find slightly unconvincing, despite its very real success.

Most development has followed van der Weide and Deconinck [39], who proposed to extend the analysis carried out for the scalar case to systems of equations by devising matrix-valued formulas that were equivalent in the sense of reducing back correctly to the scalar case, and possessed other useful properties. However, it has proved difficult to give these formulas a clear independent interpretation other than as a superposition of plane wave solutions. A different approach was taken by Mesaros [24] who divided the residual into parts that represent elliptic and hyperbolic effects. Only the hyperbolic part of the residual is treated by an extension of the scalar method. The elliptic part of the residual is reduced by a least squares minimization that is direction-free. This method is the only "upwind-inspired" method that I know of to possess an accurate incompressible limit. Nishikawa et al. [26,30] extended it to third-order and Nishikawa and van Leer [27] made it the basis for a transonic airfoil code that achieved textbook multigrid. Unfortunately, there were difficulties going to three dimensions or time-dependent flows. The approach to be described here is free from those limitations. It takes a different approach to distinguishing varieties of multidimensional behavior, based on distinguishing advective and non-advective behavior, rather than hyperbolic and elliptic behavior.

In higher dimensions, almost all sets of conservation laws contain advective terms $(\mathbf{v} \cdot \nabla)\mathbf{u}$ that describe how some quantity is transferred from place to place. In the absence of other terms, the current local value of that quantity would have been found, at a previous time, somewhere along the particle path. This aspect of the problem deals with the propagation of information by the moving medium. It is pure interpolation and because the most relevant data is that closest to the particle path, upwinding is appropriate, and there is a clear basis for limiting, because previous maximum and minimum values are available.

All other terms involve propagation of information through the medium. Acoustic waves are the most obvious example. They are nonmaterial, and in the absence of advection the domain of dependence is symmetrical. Upwinding is not necessary because data in every direction is relevant. Because wave focussing may occur, there are no bounds on the behavior of the solution, and therefore no clear and rigorous basis for limiting. Although the underlying motive for all “upwind” methods is purportedly to enforce the correct propagation of informations, the current implementations achieve this only for one-dimensional problems.

The new methods described here are still finite-volume methods, but the solution is taken to be continuous across cell boundaries, so that no Riemann problems arise. The fluxes are instead independent variables obeying their own rules of evolution, and so the name “Active Flux” method has been adopted. This almost automatically improves the accuracy in comparison to a regular finite-volume scheme because each cell is associated with more degrees of freedom. This is also true of discontinuous Galerkin methods, but to achieve any particular order of accuracy, the Active Flux methods need less storage than discontinuous Galerkin methods and do not rely on superconvergent behavior. It is in the evolution of the fluxes that physics is accounted for. We will focus on third-order methods as being (in my opinion) especially appropriate within the aerospace industry, although higher-order is
certainly possible. But at third order the method remains rather simple, and is fully discrete and explicit, allowing a time step up to the maximum permitted by the physics.

The present paper is the prelude to a series that will confront the issue of correctly propagating different kinds of information in higher dimensions by dividing the Euler equations into their advective and acoustic components. This is a form of operator splitting but at the linear level it incurs no error because in fact these operators commute. It turns out that even for nonlinear problems there are very few interaction terms, and these do not affect the accuracy of the method.

The beginning of my most recent efforts to develop a method without discontinuities was the investigation of a method due to Karabasov and Goloviznin [18] which also features “active” fluxes. Together with my student Timothy Eymann, we derived a family of second-order schemes of this form for one-dimensional linear advection, and came to realize [9] that the optimal, third-order, member of the family was in fact the scheme derived in 1977 by van Leer as part of his “new approach to numerical advection” [40] and denoted by him as Scheme V. An appreciation of van Leer’s paper, with some additional one-dimensional schemes, has been presented by Rider [31]. Several authors since van Leer have rediscovered this advection scheme, apparently independently. See Popov and Ustyugov [29], and Akoh et al. [2]. A large timestep version has been devised by Berthon et al. [4] and a semi-discrete form by Zeng [42]. Akoh et al. [2] have made a version that they have applied to the two-dimensional shallow water equations, although many details of their approach differ somewhat from ours. For example, their reconstruction is not continuous and the method does not directly extend to three dimensions. We employ a different extension of Scheme V, valid in any number of dimensions, to compute the fluxes due to advection.

The application to wave equations is similar in principle to Scheme V, in that it is based on the exact solution to a linear model problem. For the acoustic system this is an extension of Poisson’s integral solution to the initial-value problem for the scalar wave equation [10]. For the scalar wave equation itself, a related method was developed independently by Hagstrom [3,15] and co-workers. The extension presented here to nonlinear acoustics in the presence of vorticity was given by Fan and Roe [12], where it was shown to give far more isotropic results than a DG method of the same formal accuracy. Nonlinear effects were included by careful choice of local values for the coefficients in the equations.

This paper will focus on presenting motivational material and outlining the general approach. Fuller details of the Active Flux method can be found in my students Ph.D. theses [11,23], and will appear in papers that are under preparation.

3 Re-examining One-Dimensional Methods

3.1 Advection in One Dimension

van Leer’s Scheme V Let us stay for the moment with the one-dimensional advection equation, \( \partial_t u + a \partial_x u = 0 \), which has certainly served as a testbed for the development of many algorithms, including the Projection-Evolution-Reconstruction algorithms introduced by van Leer in 1977 [40]. The general idea of these methods is to consider only data in the class that can be described by some given set of basis functions \( b_j(\xi) \) where \( \xi = \frac{2(x-x_\bar{c})}{\Delta} \in [-1, 1] \) is a local coordinate in each cell, and \( a_{j,k} \) is the coefficient of \( b_j \) in the cell \( k \). The basis functions are restricted by requiring that each of them vanishes for values of the argument outside the range \([-1, 1]\), and set
Fig. 2 Forming the flux for linear advection according to Scheme V by tracing characteristics. The flux through the right boundary is

\[ \Delta t \xi F' E' \]

which allows the exact solution to be written immediately as

\[ u(\xi, 0) = \sum_{k=1}^{k=k_{\text{max}}} \sum_{j=1}^{j=j_{\text{max}}} a_{j,k} b_j(\xi), \]

where

\[ b_j(\xi) = \chi(\xi) \]

Each of these choices leads to linear reconstructions that are not \( C^0 \) continuous, and therefore must be supplemented by a Riemann solver.

In [40] van Leer considered also the effect of adding quadratic variation to give a reconstruction

\[ u_j(\xi) = \frac{3(1-\xi^2)}{2} \bar{u}_j - \frac{(1+3\xi)(1-\xi)}{4} u_{j-\frac{1}{2}} - \frac{(1-3\xi)(1+\xi)}{4} u_{j+\frac{1}{2}}, \]

where \( \bar{u} \) is the mean value in the cell, and \( u_L = u_{j-1/2}, u_R = u_{j+1/2} \) are values at the left and right cell boundaries. Because these boundary values \( u_{j+1/2} \) are shared by both of the adjacent cells \( j, j+1 \) this reconstruction is continuous. Therefore the required storage is two units per cell, exactly the same as MUSCL. Because the boundary values are shared, however, quadratic reconstruction is possible, and third-order accuracy can be achieved. He called this Scheme V. Because the reconstruction is continuous, no Riemann solver is required. The method is upwinded simply because the flux through each interface is found from the data on the upwind side of it.

Note that the reconstruction makes no use of data outside the cell, so that the accuracy of the reconstruction is independent of mesh geometry. This is a property that can and will be
preserved in higher dimensions. Taking $a > 0$, and defining the Courant number $\nu = \frac{a \Delta t}{\Delta x}$, the flux leaving through the right-hand boundary is

$$\bar{f}_{j+\frac{1}{2}} = a \int_{1-\nu}^{1} u(\xi) \, d\xi = a \left[ v(3 - 2\nu)\bar{u}_j - v(1 - \nu)u_{j-\frac{1}{2}} + (1 - \nu)^2 u_{j+\frac{1}{2}} \right],$$

which is the average value in the shaded region of Fig. 2. It will be useful to observe that of course exactly the same result is obtained by numerical integration (See again Fig. 2), first calculating

$$f_{F'} = f(1, \frac{1}{2} \Delta t) = a \left[ \frac{3}{2} v(2 - \nu)\bar{u}_j - \frac{1}{4} v(4 - 3\nu)u_{j-\frac{1}{2}} + \frac{1}{4} (2 - v)(2 - 3\nu)u_{j+\frac{1}{2}} \right],$$

$$f_{E'} = f(1, \Delta t) = a \left[ 6v(1 - \nu)\bar{u}_j - v(2 - 3\nu)u_{j-\frac{1}{2}} + (1 - v)(1 - 3\nu)u_{j+\frac{1}{2}} \right],$$

and then using these to integrate the flux by Simpson's Rule.

$$\bar{f}_{j+\frac{1}{2}} = \frac{1}{6} [ f_R + 4f_F + f_E ],$$

because $u_{F'} = u_F$, etc. This version is helpful in the nonlinear case because then the exact integrals become cumbersome.

Once we have the fluxes, we make a conventional finite-volume update

$$\bar{u}_{j}^{n+1} = \bar{u}_{j}^{n} - \frac{\Delta t}{x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}} \left[ \bar{f}_{j+\frac{1}{2}} - \bar{f}_{j-\frac{1}{2}} \right],$$

(9)

(10)

(11)

with

$$G_V = \left[ \begin{array}{cc} 1 - (1 - e^{-i\theta}) v(3 - 2\nu) & -(1 - e^{-i\theta}) v(1 - \nu) \left( 1 - v - v e^{-i\theta} \right) \\ 6 v (1 - \nu) & (1 - \nu) (1 - 3\nu) - v (2 - 3\nu) e^{-i\theta} \end{array} \right].$$

Because we now deal with a $2 \times 2$ system rather than a scalar equation this matrix has two eigenvalues $g_{1,2}$. In many situations where more than one root arises, one of these roots would
be regarded as spurious, and was so described in [40]. This is the case with leapfrog schemes, but here, the additional root exists because we have two independent degrees of freedom for each cell, which means that we can represent frequencies beyond the Nyquist limit of $\theta = \pi$.\(^3\)

The two roots each represent legitimate solutions, in the separate ranges $[0, \pi]$ and $[\pi, 2\pi]$. Thus the active flux schemes have twice the resolving power of regular schemes. Indeed this is also true of Discontinuous Galerkin schemes. It is tedious but routine to confirm that neither of the eigenvalues is outside the unit circle if $\nu \leq 1.0$. The amplification factor for a signal proportional to $e^{i\theta}$ is, for small $\theta$

$$|g| = 1 - \frac{1}{72} \nu(1 - \nu)(1 - \nu + \nu^2)\theta^4 + O(\theta^6)$$ (12)

and the largest absolute value for the coefficient of $\theta^4$ in the range $\nu \in [0, 1]$ is 1/384. For comparison, the standard third-order upwind finite-difference method has

$$|g| = 1 - \frac{1}{24} \nu(1 - \nu^2)(2 - \nu)\theta^4 + O(\theta^6)$$ (13)

and the corresponding coefficient has a maximum absolute value of 3/128.

The semi-discrete form of Scheme V has also been rediscovered [42]. It can be written as

$$\partial_t \left( \bar{u}_j \frac{u_{j+\frac{1}{2}}}{u_{j+\frac{1}{2}}} \right) = \frac{a}{\Delta x} \begin{bmatrix} 0 & e^{-i\theta} - 1 \\ 6 & -4 - 2e^{-i\theta} \end{bmatrix} \begin{bmatrix} \bar{u}_j \\ u_{j+\frac{1}{2}} \end{bmatrix}. \tag{14}$$

The matrix $G_V'$ here is obtained as $\lim_{\nu \to 0} \partial_t G$.

The Discontinuous Galerkin method

The dispersion analysis for discontinuous Galerkin methods is given by Sherwin [38], Cheng and Shu [5], and by Kiridonova and Qin [19]. We will not repeat this analysis but merely remind the reader of the main idea in the simplest case. Within each cell $j$ the solution is assumed to be piecewise linear, and may be written in terms of two linear basis functions $\phi_{\pm} = 1 \pm x$

$$u_j(\xi) = (1 - \xi)u_j^- + (1 + \xi)u_j^+ = \phi^- u_j^- + \phi^+ u_j^+. \tag{15}$$

The $u_j^\pm$ are values just inside cell $j$ at $\xi = \pm 1$. They are not shared with the adjacent cells, so there are again two degrees of freedom per cell. The governing equation is expressed in a weak form as

$$\int_{-\infty}^{\infty} \psi(\xi)((1 - \xi)\dot{u}_j^- + (1 + \xi)\dot{u}_j^+) + a\partial_x u(\xi) d\xi = 0, \tag{16}$$

where dots denote time derivatives and $\psi(\xi)$ is one of two test functions. It can be shown that the truncation error is minimised by taking the test functions equal to the basis functions, so that $\psi(\xi) = \phi_\xi$. After integrating the spatial derivative terms by parts we have the two equations,

$$\int_{-1}^{1} (\phi^-)^2 d\xi \dot{u}_j^- + \int_{-1}^{1} \phi^+ \phi^- d\xi \dot{u}_j^+ - f(-1) + \frac{a}{2} \left( u_j^- + u_j^+ \right) = 0, \tag{17}$$

$$\int_{-1}^{1} \phi^+ \phi^- d\xi \dot{u}_j^- + \int_{-1}^{1} (\phi^+)^2 d\xi \dot{u}_j^+ + f(1) - \frac{a}{2} \left( u_j^- + u_j^+ \right) = 0. \tag{18}$$

where $f(u) = au$ is the flux function. If these equations are integrated we find

$$\Delta x_j \left( \frac{1}{3} \dot{u}_j^- + \frac{1}{6} \dot{u}_j^+ \right) - f(-1) - \frac{a}{2} \left( u_j^+ + u_j^- \right) = 0, \tag{19}$$

\(^3\) The method is applicable even at boundaries, so no boundary conditions need be considered in the analysis.
\[ \Delta x_j \left( \frac{1}{6} \dot{u}_j^+ + \frac{1}{3} \dot{u}_j^- \right) + f(1) + \frac{a}{2} \left( u_j^+ + u_j^- \right) = 0. \]  

(20)

The interface numerical fluxes \( f(\pm \frac{1}{2}) \) are assigned, as in the finite-volume method, by solving a Riemann problem to obtain the upwind flux, \(^4\)

\[ \begin{align*}
    f(-1) &= au_{j-1}^+, \\
    f(1) &= au_j^+,
\end{align*} \]

leading to update formulas that can be expressed in the matrix form:

\[
\partial_t \begin{pmatrix} u^- \\ u^+ \end{pmatrix} = \frac{a}{\Delta x} \begin{pmatrix}
-3 & -1 + 4e^{-i\theta} \\
3 & -1 - 2e^{-i\theta}
\end{pmatrix} \begin{pmatrix} u^- \\ u^+ \end{pmatrix}.
\]

(21)

Although derived quite differently, this is actually identical to the semi-discrete version of Scheme V described by Eq. 14. The matrix \( G'_{DG} \) appearing in this equation has the same trace and determinant as \( G'_V \) appearing in (14) so the matrices are similar, in the sense of having the same eigenvalues. The two schemes are the same, apart from being expressed in different bases. In fact

\[
G'_{DG} = TG'_V T^{-1}
\]

(22)

where \( T = \begin{pmatrix} 2 & -1 \\ 0 & 1 \end{pmatrix} \) is the matrix that transforms from the Scheme V basis of (mean value, right state) to the DG basis of (left state, right state). As with Scheme V, the second eigenvalue need not be thought of as spurious, but as giving results in the frequency range \( \theta \in [\pi, 2\pi] \).

### 3.2 Comparing AF and DG1

We will compare the dispersion relationships of the semi-discrete schemes with those of Scheme V at various Courant numbers. To compare the semi-discrete and fully-discrete forms, we will assume perfect time integration for the semi-discrete form,

\[
u(t) = e^{\frac{at}{\Delta x}} G' u(0),
\]

so that at \( t = \Delta x/a \), equivalent to one timestep at a CFL number of unity, the amplification factor is just \( G_{sd} = e^{\Re(g)} \), where \( g \) is the appropriate eigenvalue of \( G \). A comparable measure for fully-discrete schemes is to calculate an amplification factor

\[
G_{fd} = |g(\nu)|^{\frac{1}{\nu}},
\]

(23)

realizing that the number of timesteps required to reach \( t = \Delta x/a \) would be \( 1/\nu \). The comparison is valid even if \( 1/\nu \) is not an integer.

The results are compared in Figs. 3 and 4. There is a considerable reduction in damping from use of the fully discrete scheme, especially in the range of frequencies \( \theta \in [\pi/4, \pi/2] \). The phase errors are also improved, especially at very high frequencies. The phase of Scheme V is exact at \( \nu = 0.5 \). This is a general result for schemes with upwind-biased stencils, demonstrated by van Leer [40]. However, this property is lost in the semi-discrete version, as also of course is exactness for \( \nu = 1.0 \), even with exact time-stepping.

On this basis we can make an admittedly naive comparison between Scheme V and DG1, which are both third-order methods.

1. Comparable results are obtained from Scheme V and DG1 when Scheme V uses half the number of cells. This gives Scheme V the superior effectiveness by a factor of 2.

\(^4\) Other choices, such as the Rusanov flux, are possible but inferior.
Fig. 3  The amplification factor for Scheme V at Courant numbers $\nu = 3/4$ (longdash), 1/2 (dash), and 1/4 (dashdot). The solid line is the limiting case, CFL = 0, which is also the amplification factor for the DG1 method.

Fig. 4  The phase speed for Scheme V at Courant numbers $\nu = 3/4$ (longdash), 1/2 (dash, also exact), and 1/4 (dashdot). The solid line is the limiting result for CFL = 0, which is the phase error for the DG1 method and all other semidiscrete forms.

2. This allows time steps that are twice as large, increasing the effectiveness to a factor of 4.
3. Since RK3 requires three stages to complete a timestep, the factor becomes 12.
4. In two or three dimensions, the factor becomes respectively 24 and 48.

Without taking this analysis too seriously, it does suggest that the Active Flux method has the potential to outperform DG1 by quite a large factor. Moreover, it appears that DG1 is only second-order accurate for the full Euler equations, whereas the Active Flux method (which is the generalization of Scheme V) has so far proved to retain third order accuracy. Any comparison should then perhaps be made with the substantially more expensive DG2.

**DG with inexact time-stepping**  In practice the DG method will need to employ some time integrator that is less than exact, and usually the time integration is performed by one of
Fig. 5 Forming the flux for Burgers’ equation by tracing characteristics. The flux through the right boundary is \( \frac{1}{12}(a_R^2 + 4u_F^2 + u_E^2) \)

the Runge–Kutta methods. The lowest-order time integrator that will preserve third-order accuracy is of course RK3, which provides stability to the DG1 discretization only up to \( \nu = 0.409 \) so that the numbers in the list above should be multiplied by about 2.5. More accurate timestepping, or the use of Strong-Stability-Preserving timemarching, will reduce the available Courant numbers further.

**Remark on Superconvergence** It is well known that discontinuous Galerkin methods employing polynomial reconstructions of order \( p \) frequently exhibit an order of accuracy \( 2p + 1 \) instead of the \( p + 1 \) that typifies finite-volume methods. For example, the DG1 method might be expected to be second-order accuracy, but turns out to be third-order. The fact that it is the semi-discrete limit of Scheme V goes a long way to explain this. Although Scheme V only stores two degrees of freedom per element, each of them is actually shared between two elements and allows the reconstruction of a quadratic rather than a linear function. If DG1 is regarded in the same way, with the outflow value regarded as also being, for the purpose of reconstruction, the inflow value of the downstream neighbor, a third-order quadratic reconstruction can be performed. This conclusion applies at all orders of accuracy [36] and the insight enables “superconvergent” Galerkin methods for linear advection to be derived with any accuracy using continuous reconstruction. Observe, however, that this continuous reconstruction is completely different from what is usually described as continuous Galerkin.

### 3.3 Burgers’ Equation

A natural next step from linear advection is Burgers’ equation.

\[
\partial_t u + u \partial_x u = 0.
\]  

(24)

There is of course no unique way to extend the solution of a linear problem to a nonlinear problem, but there is a very simple modification (see Fig. 5) that does extend to more general problems. As for the linear advection equation, we first find the solution on the outflow boundary \( x = x_R \) by tracing back the characteristics from the quadrature points \( E', F' \). If this characteristic issues from the point \( x = x_R + x^* \) and carries the value \( u = u^* \) then

\[
x^* = -\tau u^* = -\tau(u_R + x^* \partial_x u)
\]
so that
\[ x^* = -\frac{u_R \tau}{1 + \tau \partial_x u} \simeq -u_R \tau (1 - \tau \partial_x u). \] (25)

For linear advection, the characteristic originates at \( x^* = -a \tau \), so we are replacing the constant advection speed \( a \) with a local speed \( u_R (1 - \tau \partial_x u) \), where \( \tau \) is either \( \Delta t \) or \( \frac{1}{2} \Delta t \). This simple first-order correction to the wavespeed carries over to the nonlinear wave equation [12] and forms the basis for our treatment of the Euler equations.

Figure 6 shows a typical solution to Burgers’ equation using this method. The initial profile is
\[ u(x) = 0.05 + 0.95 \exp(-50(x - 0.5)^2) \]
which steepens into a shock at \( t = 0.173 \). The solution is found here at \( t = 0.150 \) just before shock formation. Table 1 shows that a convergence rate close to 3.0 is achieved by incorporating the nonlinear correction. Further from the shock, the convergence is much more rapid. The treatment after a shockwave has formed will be part of a more detailed paper.

4 The Multidimensional Strategy

4.1 The Overall Plan

The purpose of this section is to set out a strategy for applying the ideas of Scheme V to multidimensional nonlinear systems displaying both advective and acoustic behavior. This will be explained at third order for two-dimensional problems on triangular grids. Non-simplicial grids can also be handled but will not be discussed here. Arbitrary order is attainable straightforwardly for linear problems, but nonlinear terms have to be handled with more care for orders greater than three.
Table 1  Accuracy of Burgers’ Solution

| Grid | No wavespeed correction | With wavespeed correction |
|------|--------------------------|---------------------------|
|      | L2 error | L2 order | L2 error | L2 order |
| 80   | 5.4047E−04 | – | 1.2746E−04 | – |
| 160  | 5.1270E−04 | 0.08 | 3.0595E−05 | 2.06 |
| 320  | 1.2928E−04 | 1.99 | 5.2251E−06 | 2.55 |
| 640  | 3.2350E−05 | 2.00 | 7.6770E−07 | 2.77 |
| 1280 | 8.0974E−06 | 2.00 | 1.0352E−07 | 2.89 |
| 2560 | 2.0264E−06 | 2.00 | 1.3428E−08 | 2.95 |
| 5120 | 5.0689E−07 | 2.00 | 1.7308E−09 | 2.96 |

The essence of the method resides in applying appropriate numerics to appropriate physics, which means treating advective and acoustic behavior differently. The data at time \( n \Delta t \), that is \( u^n \), is updated to \( u^* \) by considering pure inertial evolution with all pressure gradients turned off. Then the partial solution at \( u^* \) is updated to \( u^{n+1} \) by considering only acoustic changes. Both processes are treated by novel methods that combine low memory, large timesteps, and high accuracy with continuous reconstructions.

The solution is represented in the usual finite-volume manner by giving the average values of the conserved variables in every cell, together with some other information on the cell boundaries. These will be used to compute the fluxes, although the fluxes themselves are not suitable to store because they are usually ambiguous in specifying the state. We have used the primitive variables at the vertices of simplex elements and at the midpoints of their edges to define quadratic distributions. Such a representation reduces storage because much data is now shared between elements, and therefore reduces the size of any linear algebra problems that might arise. One timestep consists of the following three stages;

1. The data on the boundary of each element is updated by some scheme that respects the physics. This means that the advective terms are applied first and upwinded and the acoustic terms are applied to the partial update without upwinding. The two stages are treated in more detail below.
2. All boundary data is converted to fluxes, and the cell average is updated by integrating them. This raises the order of accuracy by one, compared to the boundary values.
3. The cell averages and boundary values are reconciled. We currently have two ways to do this. One is to add a bubble function to the primitive variables, to produce agreement between the cell averages when the primitive variables are integrated over the cell. Another is to measure the discrepancy between those averages by making corrections to the primitive variables.

For linear problems the first step is simple, because the advective and acoustic operators commute. We can carry them out independently, in either order. In the nonlinear case, most of the nonlinear terms can be absorbed into the coefficients of a linear problem, as with Burgers’ equation, but a few terms need careful analysis. This will be described in a paper that is being prepared.

---

5 This observation is in itself an argument against discontinuous reconstructions.
4.2 A New Approach to Multidimensional Advection

Van Leers approach to advection has often been generalized to multidimensional linear advection $\partial_t u + a \cdot \nabla u = 0$. It is easily accomplished in principle by representing the data as a sum of local basis functions within each cell. Each of these functions is translated by $a \Delta t$ and the new solution is projected onto the basis functions. The process has much in common with the “remap” phase of a Lagrangian–Eulerian calculation [31] and is usually carried out in the context of discontinuous reconstructions.

To deal with nonlinearity we take a different view, and regard the advective stage of the Euler solution as one in which particles preserve their momentum until they collide. They obey the pressureless Euler equations
\begin{equation}
\partial_t u + u \cdot \nabla u = 0,
\end{equation}
which is a generalization of Burgers’ equation to $n$ dimensions and has a similar exact, but implicit, solution, which is $u = U(x - ut)$, for any arbitrary function $U : \mathbb{R}^n \rightarrow \mathbb{R}^n$. so that in a small neighborhood
\begin{equation}
x^* = -\tau u^* = \tau (u + x^* \nabla u)
\end{equation}
and
\begin{equation}
x^* = (I - \tau \nabla u)^{-1} \tau u
\end{equation}
This is the correct generalization of (25) However, (26) is not in conservation form, so cannot be completely treated with the Active Flux method, although the second-order solution $u(0, \tau) = u(x^*, 0)$ is good enough to update the point values. Then, the conservative form of the momentum equations is employed for the final update
\begin{equation}
\partial_t (\rho u) + \nabla \cdot (\rho u \otimes u) = 0.
\end{equation}
This describes pure inertial motion provided that the continuity equation,
\begin{equation}
\partial_t \rho + \nabla \cdot (\rho u) = 0,
\end{equation}
is also satisfied, and this will require also the point values of $\rho$. These follow from continuity. Along any pathtube we have $\rho(t)S(t) = \text{const}$, where $S(t)$ is the area of a patch of fluid. Each point of the patch moves with a constant speed $u_0$, so the mapping from location at $t = 0$ to $t = \tau$ is
\begin{equation}
x(\tau) = x(0) + u(x(0))\tau
\end{equation}
and
\begin{equation}
J = \frac{\partial x(\tau)}{\partial x(0)} = \begin{pmatrix}
1 + \tau \frac{\partial u}{\partial x} & \tau \frac{\partial u}{\partial y} \\
\tau \frac{\partial v}{\partial x} & 1 + \tau \frac{\partial v}{\partial y}
\end{pmatrix}.
\end{equation}
It follows that the density is given by
\begin{equation}
\rho(0, \tau) = \frac{\rho(x^*, 0)}{\det J(x^*, 0)}.
\end{equation}
This result is exact for the pressureless Euler equations until fluid paths collide, although because we are using this result embedded into the full Euler equations, the pressure terms there will prevent collision.

Numerical implementation is sketched in Fig. 7, which shows three triangular elements and the piecewise parabolic distributions inside them. On each interface between triangles, characteristics are traced back from the quadrature points needed for Simpsons Rule, in order
to locate the points $x^*$. Care must be taken to avoid double counting when streamlines almost coincide with an edge.

### 4.3 A New Approach to Multidimensional Wave Propagation

The scalar wave equation is $\partial_t \phi = c^2 \nabla^2 \phi$, for which the classical Poisson solution [7] gives the general solution to the initial-value problem. However, this is not the form in which wave behavior appears in the Euler equations. Instead we find (at the linear level) the first-order system

\begin{align}
\partial_t p + \nabla \cdot \mathbf{u} &= 0 \quad (33) \\
\partial_t \mathbf{u} + \nabla p &= 0 \quad (34)
\end{align}

It is easy to show that the pressure always obeys the classical solution, and so can be found from (35), but the velocity does not, unless the flow is irrotational. If the flow initially contains vorticity, then we have simply $\partial_t \nabla \times \mathbf{u} = 0$ and vorticity is preserved. Because the derivation is quite lengthy, it is not given here, but the general solution to the initial-value problem is

\begin{align}
p(x, t) = & M_{ct} p(x, 0) - ct M_{ct} \nabla \cdot \mathbf{u} + \int_0^{ct} t M_{ct} \nabla^2 p(x, t) \, dt \\
\mathbf{u}(x, t) = & M_{ct} \mathbf{u}(x, 0) - ct M_{ct} \nabla p + \int_0^{ct} t M_{ct} \nabla (\nabla \cdot \mathbf{u}(x, t)) \, dt \quad (36)
\end{align}

Just as with van Leer’s much earlier treatment of linear advection, a numerical method follows from expressing the initial data in any finite-element basis and then evaluating these integrals to predict the (active) fluxes on the boundaries of the control volumes. For piecewise polynomial data the integrals have simple closed forms. Figure 8 shows how the residual at each node is the sum of integrals over segments of circular discs in each neighboring element. This can be efficiently coded as a loop over the elements.

Since a necessary condition for stability is that the numerical domain of dependence must contain the physical domain of dependence, it follows that the time step must be restricted.
so that no Mach disc escapes from the union of elements to which the corresponding node belongs. It has been found in practice that time steps extremely close to that limit may be used on unstructured grids.

Tests were carried out on the classical problem of an initial Gaussian pulse in pressure. The nonlinear \( p \)-system was chosen, so that

\[
\begin{align*}
\partial_t \rho + \nabla \cdot u &= 0, \\
\partial_t u + \nabla p(\rho) &= 0,
\end{align*}
\]

with \( p(\rho) = \rho^\gamma \). The Active Flux method was found to offer much better resolution than DG1 (which is still third-order for this problem). Figure 9 shows that the Active Flux method does a much better job of capturing the peak value, and has only about a quarter of the vertical scatter.

This approach may be compared with the evolution Galerkin method developed by Lucakova-Medvidova and Morton [22], which also uses the wave equation as a jumping-off point, but retains a discontinuous representation and makes an approximate integration of the bicharacteristic equations. This only achieves second-order accuracy, on an enlarged stencil.

**Fig. 8** Each node is updated by integration over the Mach disc enclosing it. \( a \) Cells contributing to node update, \( b \) cells contributing to edge update

**Fig. 9** Two snapshots of a spreading circular wave close to its peak. Red is DG1, blue is AF (Color figure online)
Fig. 10  Accuracy of the density in a translating vortex for four compressible flow methods, (left) as a function of the degrees of freedom, and (right) as a function of the required work units.

5 The Euler Equations

As already mentioned, the plan is to apply operator splitting to the advective and acoustic aspects of the flow, despite the nonlinear nature of the problem. There are several options for this, and so again the details are reserved for a future publication. However, it proves possible to overcome the nonlinearities by absorbing them into the coefficients of a local linear problem.

Tests have been carried out for the translating vortex problem, which was first suggested by Chi-Wang Shu and has proved a most valuable resource in high-order code development. The specific example used was taken from the High-Order Workshop [41].

Figures 10 and 11 show the errors in density and velocity from four different schemes. Included for comparison are the results from a second-order MUSCL scheme. However, this method was implemented only on a simple square mesh and does not have to deal with geometrical terms. This code does feature a nonlinear limiter, although this was probably not activated by this mildly nonlinear problem. The results for DG1 and DG2 were obtained from a suite of programs written by Professor Chris Fidkowski. These codes were intended for educational purposes to display results from a variety of methods. The results for Active Flux come from research codes written by Doreen Fan and Brad Maeng as part of their work toward their doctorates.

On the left, the methods are compared for given degrees of freedom, by plotting against the effective mesh size \( h = (\text{d.o.f.})^{-1/2} \). The MUSCL and DG1 results are clearly second-order accurate, and quite surprisingly similar. AF and DG2 are both clearly third-order accurate, with some superiority attaching to DG2.

A comparison in terms of work units presented difficulty since none of the codes were written with efficiency as a major objective. A crude handicapping method was adopted, in which the degrees of freedom were multiplied by the number of updates needed to advance...
In fact, there is much scope for argument about the fairness of these handicaps, but the differences revealed are too large to be ignored. For this problem, and this range of grid sizes, DG1 and DG2 turn to be quite comparable in accuracy, despite one being second order and the other one third-order. But the Active Flux method is clearly superior to both. For given work units the accuracy either in velocity or density is about one and a half orders of magnitude better. For given accuracy the required work is lower by one order of magnitude. Significantly, the Active Flux method could be run up to a CFL number of 1.0, as was the case with our model problems.

6 Discussion

The striking success of high-resolution Finite-Volume methods, and discontinuous Galerkin methods, both of which are based on discontinuous representations of the solution, has lead the scientific computing community to think that such representations provide a necessary key to success in the numerical solution of hyperbolic problems. This seems doubtful, because it has long been realized that there are numerous theoretical objections, beginning with the recognition by Godunov himself that carrying this philosophy to its logical conclusion was a futile exercise. Other issues have been the poor representation of discontinuities lying oblique to the grid, especially shear and contact discontinuities, and an excess of dissipation at very low speeds. I attribute these shortcomings to the exclusive use of one-dimensional arguments in attempting to inject physics into the algorithms.
Some authors (myself included) have attempted to introduce directions that are determined by the solution rather than by the grid, but I do not now find any of them convincing; they compensate for bad physics rather than replacing it. The failure at low Mach numbers should come as no surprise when we realize that one-dimensional unsteady incompressible flows can only exist with constant velocity, and this makes the idea of a Riemann problem rather absurd. Although numerous empirical extensions have been proposed, they do not compete with codes specifically designed for incompressible flow.

The Active Flux method whose principles are described here avoids both of these issues by not introducing unjustified discontinuities, and also by insisting that every type of behavior in a multidimensional compressible flow shall be represented through an appropriate discrete operator. That is, after all, precisely the philosophy behind upwinding as it is very successfully applied to one-dimensional flow. However, in higher dimensions there are a greater number of distinct behaviors to be provided with distinct treatments. In consequence, many of the techniques developed for compressible flow in recent decades do not apply, and new techniques have had to be invented. So far, this has proved possible, and the new ideas have proved to be simple and accurate. Much remains to be done, but comparison between the new and the old is extremely encouraging.

In various tests, the method based on continuous reconstructions has proved more cost-effective by least an order of magnitude. Good progress is being made at the moment on a steady-state version, and on limiting methods. The principle challenge ahead is the extension to Navier–Stokes, and the intention is to adopt the hyperbolic versions of Navier–Stokes recently put forward by Nishikawa [25] and by Peshkov and Romenski [28]. This will involve considering, as in solid mechanics, both compressive and torsional waves, and it is encouraging that these will also commute, and may be treated independently.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Abgrall, R.: A review of residual distribution schemes for hyperbolic and parabolic problems: the July 2010 state of the art. Commun. Comput. Phys. 11(04), 1043–1080 (2012)
2. Akoh, R., Ji, S., Xiao, F.: A multi-moment finite volume formulation for shallow water equations on unstructured mesh. J. Comput. Phys. 229(12), 4567–4590 (2010)
3. Alpert, B., Greengard, L., Hagstrom, T.: An integral evolution formula for the wave equation. J. Comput. Phys. 16(22), 536–543 (2000)
4. Berthon, C., Sarazin, C., Turpault, R.: Space-time generalized Riemann problem solvers of order k for linear advection with unrestricted time step. J. Sci. Comput. 55(2), 268–308 (2013)
5. Cockburn, B., Shu, C.-W.: RungeKutta discontinuous Galerkin methods for convection-dominated problems. J. Sci. Comput. 16(3), 173–261 (2001)
6. Colella, P.: Multidimensional upwind methods for hyperbolic conservation laws. J. Comput. Phys. 87(1), 171–200 (1990)
7. Courant, R., Hilbert, D.: Methods of Mathematical Physics, vol. II. Interscience, New York (1962)
8. Deconinck, H., Pailiere, H., Struijs, R., Roe, P.L.: Multidimensional upwind schemes based on fluctuation-splitting for systems of conservation laws. Computational Mechanics 11(5–6), 323–340 (1993)
9. Eymann, T.A., Roe, P.L.: Active flux schemes. In: 49th AIAA Aerospace Sciences Meeting (2011)
10. Eymann, T.A., Roe, P.L.: Multidimensional active flux schemes. In: 21st AIAA Computational Fluid Dynamics Conference (2013)
11. Fan, D.: On the acoustic component of active flux schemes for nonlinear hyperbolic conservation laws. Ph.D. thesis, Aerospace Engineering, University of Michigan (2017)
12. Fan, D., Roe, P.L.: Investigations of a new scheme for wave propagation. In: 22nd AIAA Computational Fluid Dynamics Conference (2015)
13. Godunov, S.K.: A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. Mat. Sb. 89(3), 271–306 (1959)
14. Godunov, S.K.: Reminiscences about difference schemes. J. Comput. Phys. 153(1), 6–25 (1999)
15. Hagstrom, T.: High-resolution difference methods with exact evolution for multidimensional waves. Appl. Numer. Math. 93, 114–122 (2015)
16. Hughes, T.J., Mallet, M.: A new finite element formulation for computational fluid dynamics: III. The generalized streamline operator for multidimensional advective–diffusive systems. Comput. Methods Appl. Mech. Eng. 58, 305–328 (1986)
17. Hughes, T.J., Scovazzi, G., Tezduyar, T.E.: Stabilized methods for compressible flows. J. Sci. Comput. 43(3), 343–368 (2010)
18. Karabasov, S.A., Goloviznin, V.M.: Compact accurately boundary-adjusting high-resolution technique for fluid dynamics. J. Comput. Phys. 228(19), 7426–7451 (2009)
19. Krivodonova, L., Qin, R.: An analysis of the spectrum of the discontinuous Galerkin method. Appl. Numer. Math. 64, 1–18 (2013)
20. Lax, D., Liu, X.-D.: Solution of two-dimensional Riemann problems of gas dynamics by positive schemes. SIAM J. Sci. Comput. 19(2), 319–340 (1998)
21. Leveque, R.J.: High-resolution conservative algorithms for advection in incompressible flow. SIAM J. Numer. Anal. 33(2), 627–665 (1996)
22. Lukacova-Medvidova, M., Morton, K.W., Warnecke, G.: Finite volume evolution Galerkin methods for hyperbolic systems. SIAM J. Sci. Comput. 26(1), 1–30 (2004)
23. Maeng, J.: (Brad), On the advective component of active flux schemes for nonlinear hyperbolic conservation laws. Ph.D. thesis, Aerospace Engineering, University of Michigan (2017)
24. Mesaros, L. M.: Multi-dimensional splitting schemes for the Euler equations on unstructured grids. Ph.D. thesis, Aerospace Engineering and Scientific Computing, University of Michigan (1995)
25. Nishikawa, H.: New-generation hyperbolic Navier–Stokes schemes: O (1/h) speed-up and accurate viscous/heat fluxes. In: 20th AIAA Computational Fluid Dynamics Conference (2011)
26. Nishikawa, H., Rad, M., Roe, P.L.: A third-order fluctuation-splitting scheme that preserves potential flow. In: AIAA Paper 2001–2595, 15th AIAA CFD Conference, Anaheim (2001)
27. Nishikawa, H., van Leer, B.: Optimal multigrid convergence by elliptic/hyperbolic splitting. J. Comput. Phys. 190(1), 52–63 (2003)
28. Peshkov, I., Romenski, E.: A hyperbolic model for viscous Newtonian flows. Contin. Mech. Thermodyn. 28(1–2), 85–104 (2016)
29. Popov, M.V., Ustyugov, S.D.: Piecewise parabolic method on local stencil for gasdynamic simulations. Comput. Math. Math. Phys. 47(12), 1970–1989 (2007)
30. Rad, M.: A residual distribution approach to the Euler equations that preserves potential flow. Ph. D. thesis, Aerospace Engineering and Scientific Computing, University of Michigan (2001)
31. Rider, W.J.: Reconsidering remap methods. Int. J. Numer. Methods Fluids 76(9), 587–610 (2014)
32. Roe, P.L.: Fluctuations and signals—a framework for numerical evolution problems. In: Morton, K.W., Baines, M.J. (eds.) Numerical Methods for Fluid Dynamics, vol. 11. Oxford University Press, Oxford (1982)
33. Roe, P.L.: Discrete models for the numerical analysis of time-dependent multidimensional gas dynamics. J. Comput. Phys. 63(2), 458–476 (1986)
34. Roe, P.L.: Linear Advection Schemes on Triangular Meshes. Cranfield College of Aeronautics, Cranfield (1987)
35. Roe, P.L.: Multidimensional upwinding. In: Abgrall, R., Shu, C.-W. (eds.) Handbook of Numerical Analysis, vol. 18, pp. 53–80. Elsevier, Amsterdam (2017)
36. Roe, P.L.: A simple explanation of superconvergence for discontinuous Galerkin solutions to $u_t + u_x = 0$. Commun. Comput. Phys. 21(4), 905–912 (2017)
37. Roe, P.L., Beard, L.: An improved wave model for multidimensional upwinding of the Euler equations. In: Thirteenth International Conference on Numerical Methods in Fluid Dynamics. Springer, Berlin (1993)
38. Sherwin, S.: Dispersion analysis of the continuous and discontinuous Galerkin formulations. In: Discontinuous Galerkin Methods. Springer, Berlin (2000)
39. Van der Weide, E., Deconinck, H.: Positive matrix distribution schemes for hyperbolic systems. In: Désidéri, J.A., Hirsch, C., Le Tallec, P., Pandolfi, M., Périaux, J. (eds.), ECCOMAS ’96, Computational Fluid Dynamics, pp. 747–753. Wiley, New York (1996)
40. Van Leer, B.: Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. J. Comput. Phys. 23(3), 276–299 (1977)
41. Wang, Z.J., et al.: High order CFD methods: current status and perspective. Int. J. Numer. Methods Fluids 72(8), 811–845 (2013)
42. Zeng, X.: A high-order hybrid finite difference-finite volume approach with application to inviscid compressible flow problems: a preliminary study. Comput. Fluids 98, 91–110 (2014)