Numerical Error Estimation for Nonlinear Hyperbolic PDEs via Nonlinear Error Transport[☆]

J. W. Banks*, J. A. F. Hittinger, J. M. Connors, C. S. Woodward

Center for Applied Scientific Computing,
Lawrence Livermore National Laboratory,
Livermore, California, 94551

Abstract

The estimation of discretization error in numerical simulations is a key component in the development of uncertainty quantification. In particular, there exists a need for reliable, robust estimators for finite volume and finite difference discretizations of hyperbolic partial differential equations. The approach espoused here, often called the error transport approach in the literature, is to solve an auxiliary error equation concurrently with the primal governing equation to obtain a point-wise (cell-wise) estimate of the discretization error. Nonlinear, time-dependent problems are considered. In contrast to previous work, fully nonlinear error equations are advanced, and potential benefits are identified. A systematic approach to approximate the local residual for both method-of-lines and space-time discretizations is developed. Behavior of the error estimates on problems that include weak solutions demonstrates the positive properties of nonlinear error transport.

Keywords: a posteriori error estimation, hyperbolic equations, finite volume methods, finite difference methods, weak solutions

[☆] This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and was funded by the Uncertainty Quantification Strategic Initiative Laboratory Directed Research and Development Project at LLNL under project tracking code 10-SI-013.

[*] Corresponding author

Email addresses: banks20@llnl.gov (J. W. Banks), hittinger1@llnl.gov (J. A. F. Hittinger), connors4@llnl.gov (J. M. Connors), woodward6@llnl.gov (C. S. Woodward)

1. Introduction

For decades, numerical methods of increasing sophistication have been routinely and successfully used to compute approximate solutions to time-dependent differential equations. Concurrently, methods have been developed to estimate the approximation error of such discrete solutions. These *a posteriori* error estimation techniques have been traditionally used as indicators for mesh (element) adaptivity in an attempt to reduce discretization error by employing locally finer meshes. However, the increasing emphasis on Uncertainty Quantification (UQ) for numerical simulations presents a new challenge for robust, reliable, and accurate techniques for the quantitative estimation of error in quantities of interest (QoI) derived from a simulation. In many complex multi-physics applications, it is impractical to use highly-resolved meshes to minimize the contributions of discretization error, and thus the uncertainty in computed results due to this error must be understood for proper interpretation of uncertainty analyses resulting from other input and data uncertainties.

In this paper, we are specifically interested in evolutionary methods of error estimation for finite difference method (FDM) and finite volume method (FVM) discretizations. In this error estimation approach, auxiliary evolution equations for the errors are derived, discretized, and solved in tandem with the approximate primal equations. Therefore, the computational cost is approximately twice that of solving the primal equation. The result is a discrete, signed, point-wise (cell-wise) approximation of the error that can be used subsequently to construct an estimate of the error in any number of linear or nonlinear functionals of the solution (QoIs). The effects of both error generation (destruction) and propagation are included, and, in addition, cancellation of errors in the calculation of QoIs can be obtained. In addition, when properly formulated, the nonlinear error transport approach is asymptotically correct and provides robust error estimates even for approximations of weak solutions.

Many approaches exist for global error estimation for numerical approximation of differential equations [1]. Historically, the main objective of error estimation has been to obtain a more accurate approximate solution by correcting a low-order scheme instead of resorting to a more expensive and potentially less robust higher-order discretization. Within the context of interest here, that is, providing error bounds on a discrete solution or functionals thereof, comparisons of techniques can be found in [2, 3]. Probably the most common technique for FDM and FVM is the venerable mesh-refinement-based Richardson extrapolation [4]. In this approach, solutions are computed on at least three meshes in order to fit the free parameters in an asymptotic error *ansatz*, typically an assumed power-law form. Grid convergence error estimation has the advantage that no modifications to the code are required, although care must be taken to constrain all code inputs properly to obtain self-consistent results. Richardson extrapolation may produce erroneous results if the error *ansatz* inadequately describes the true error behavior [2], which unfortunately may be the case for simulations that fail to resolve all scales in the solution. This includes cases where discontinuities are present and true weak solutions are sought. There have been attempts to consider richer *ansätze* with varying degrees of success [5].

In contrast to the mesh refinement strategy, several other classes of *single-grid* estimators exist. The finite element method (FEM) literature provides at least two broad classes of error estimators developed primarily for linear elliptic and parabolic problems: residual methods and recovery methods [6]. Typically these methods are used to produce error indicators for adaptive mesh refinement, and as such the error estimates are provided as local or global bounds in some relevant norm. Because of our focus on FDM and FVM discretizations of hyperbolic problems, we do not consider these methods further.

Another class of single-grid estimators is based on the solution of auxiliary problems. Richardson extrapolation, for instance, can be based on varying the discretization order (*cf.*, mesh refinement) [2]. Such an approach has not often been used for finite difference and finite volume methods, most likely because of the difficulty of implementing multiple, higher-order and stable discretizations. Adjoint methods [7, 8] estimate the error in a QoI from the solution of an auxiliary adjoint problem and the residual of the discrete solution. When coupled with mesh adaptivity, for a small number of QoIs, adjoint methods can be very efficient at refining only in regions to which the QoI is sensitive. However, the challenge of formulating the adjoint problem for complex multi-physics problems, the inefficiency of solving a different adjoint problem for each QoI, and the expense associated with long time integration of nonlinear hyperbolic problems can be drawbacks. With implicit residual methods [6, 9] from the FEM literature, local error estimators are developed by solving auxiliary equations for the error on a mesh element or on a patch of elements. The local errors can also be used to construct a computable estimate of some global norm of the error. However, the development of the local error equations is specific to finite element methods.

In the finite difference and finite volume literature, evolutionary error methods considered herein are often referred to as *error transport* methods. These techniques are a form of non-iterative difference or differential defect correction [1, 10]. In all cases in the error transport literature, the error equations are linearized, and the focus is typically on the approximation used to evaluate the local residual, that is, the local source (sink) of error that drives an error transport equation. Linear error transport techniques have been applied to linear and nonlinear models of advection [11, 12] and advection-diffusion [12–17] equations, subsonic and supersonic inviscid flow [11, 18, 19], and subsonic viscous flow [13, 17]. Most investigations have considered steady-state solutions [14–19], and only a few examples of time-dependent applications [11, 12] exist. A good review of previous work can be found in Hay and Visonneau [17].

In the context of this previous work, this paper has several goals. Our overarching goal is to investigate the viability of the nonlinear error transport approach as a strategy of *a posteriori* error estimation for the purposes of uncertainty quantification. We focus on time-dependent, nonlinear, hyperbolic problems that admit weak solutions and present a new, simplified procedure for approximate residual evaluation that is applicable to both method-of-lines and space-time discretizations. In addition, we argue that solving the full error equations, including any error nonlinearity, has potential advantages over the more traditional approach of linearization. We demonstrate instances where the inclusion of these terms increases the robustness of the error estimates.

The remainder of this paper is structured as follows. In Section 2, we will introduce the basic concepts of nonlinear error transport for general time-dependent partial differential equations (PDEs), and in Section 3, we will present the core ideas of our approach in contrast to previous linear error transport work. We will specialize in Section 4 to a canonical example of a non-linear hyperbolic PDE, the 1D inviscid Burgers' equation. Specific example discretizations for both method-of-lines and space-time discretizations are developed. Convergence properties of the method are demonstrated in Section 5 for strong and weak solutions, and the effects of nonlinearity in the error equation are explored in both one and two dimensions in Section 6. In Section 7, we demonstrate the application of the approach to a nonlinear system, the 1D Euler equations. Finally, we summarize our results and make conclusions in Section 8.

2. Basic Concepts

Consider an evolution equation for $u(x,t)$ of the form

$$\partial_t u + \mathcal{F}(u) = s, \tag{1}$$

where $\mathcal{F}(u)$ is some linear or nonlinear (spatial) differential operator on u and $s(x,t)$ is an inhomogeneous term. For convenience, we make the simplifying assumption that $x \in \mathbb{R}$ in order to eliminate the need to consider boundary conditions. The nature of errors introduced via boundary condition approximation is an interesting subject that will be addressed in future work. Initial conditions are given as $u(x,t=0) = g(x)$.

We now assume that we have a function $\tilde{u}(x,t)$ that approximates u but does not satisfy (1) exactly. Define the approximation error to be

$$e(x,t) = u(x,t) - \tilde{u}(x,t). \tag{2}$$

Substitution into (1) yields

$$\partial_t e + \mathcal{F}(e + \tilde{u}) = s - \partial_t \tilde{u}. \tag{3}$$

A residual is formed on the right-hand side by subtracting $\mathcal{F}(\tilde{u})$ from both sides:

$$\partial_t e + \mathcal{G}(e;\tilde{u}) = -\left(\partial_t \tilde{u} + \mathcal{F}(\tilde{u}) - s\right), \tag{4}$$

where $\mathcal{G}(e;\tilde{u}) = \mathcal{F}(e+\tilde{u}) - \mathcal{F}(\tilde{u})$. The reason for this last step becomes apparent when $\mathcal{F}$ is restricted to be a linear operator; in this case,

$$\partial_t e + \mathcal{F}(e) = -\left(\partial_t \tilde{u} + \mathcal{F}(\tilde{u}) - s\right), \tag{5}$$

that is, the error is acted on by the same operator as the primal solution and driven by the residual of that operator acting on the approximate solution. Though obfuscated somewhat by the explicit time derivative and nonlinear generalization, the relationship in (4) (or more often (5)) between error and residual are a well-known result in numerical analysis.

Equation (4) is the error equation that describes the evolution of the approximation error in space and time. In the nonlinear case, the error evolves by a *different* differential operator than the solution. In the linear error transport approach [11–19], the differential (or corresponding difference) operator on the left-hand side of (4) is linearized about $\tilde{u}$, which is a reasonable assumption so long as $|e(x,t)| \ll |u(x,t)|$. In practice, and in some very important cases, this is not valid, and such a linearization becomes questionable. We will demonstrate some of these failings in Section 6 and show that use of the nonlinear error transport equation overcomes these problems.

Through this point in the discussion, $\tilde{u}$ and e have been treated as continuous functions of x and t. We now specialize to the problem of discretization error. In general, we do not know how to solve either the original PDE (1) or the error equation (4). We can, however, discretize both and solve each approximately. There are thus three decisions to be made: the choice of discretization of the primal equation (1), the choice of discretization of the evolution operator in the error equation (4), and the choice of an evaluation technique for the residual in the error equation (4). Appropriate choices for the first two discretizations depend intimately on the nature of the continuous operator and initial conditions. Typically the same (or a similar) scheme is used for both the primal and auxiliary equations.

For smooth solutions and consistent, stable, solution-independent discretizations, we can determine the relationship of the order of the three discretizations to the rate of convergence of the error approximation. Of course, for weak solutions or nonlinear schemes, this relationship will provide only an upper bound. We take a finite-difference viewpoint with $u_i^n = u(x_i, t^n)$ and assume a smooth (infinitely differentiable) exact solution and numerical approximation; similar constructions can be made in the finite-volume framework. We also assume $\Delta x / \Delta t$ is fixed as $\Delta x, \Delta t \to 0$ in order to simplify notation. We assume that the primal scheme is order p_t in time and p_x in space, *i.e.*,

$$|e| = |\tilde{u} - u| = O(\Delta t^{p_t}) + O(\Delta x^{p_x}) = O(\Delta x^p)\,, \tag{6}$$

where $p = \min(p_t, p_x) > 0$. We write the discretization of the error equation as

$$\frac{\tilde{e}_i^{n+1} - \tilde{e}_i^n}{\Delta t} + G_i\left(\tilde{e}^n; \tilde{u}^n, \tilde{u}^{n+1}\right) = -R\left(\tilde{u}^n, \tilde{u}^{n+1}\right).$$

If the left-hand side of the discrete error equation is order q_t in time and q_x in space, and we set $q = \min(q_t, q_x) > 0$, then Taylor series expansion gives

$$\left[\partial_t \tilde{e} + \mathcal{G}\left(\tilde{e}; \tilde{u}\right)\right]_i^n = O(\tilde{e}\Delta x^q) = O\big(\Delta x^{p+q}\big)\,.$$

The $\tilde{e}$ factor appears because the Taylor series terms are derivatives of $\tilde{e}$, we have assumed that the solution is well-resolved, and $\tilde{e}$ approximates e, which by (6) is an $O(\Delta x^p)$ quantity. Assume the residual is approximated to order r_t in time and r_x in space with $r = \min(r_t, r_x) > 0$:

$$R\left(\tilde{u}^n, \tilde{u}^{n+1}\right) = \mathcal{R}(\tilde{u})|_i^n + O(\Delta x^r)\,.$$

Thus,

$$\left[\partial_t \tilde{e} + \mathcal{G}\left(\tilde{e}; \tilde{u}\right) - \mathcal{R}(\tilde{u})\right]_i^n = O\big(\Delta x^{p+q}\big) + O(\Delta x^r)\,,$$

and so the rate of convergence of the approximate error $\tilde{e}$ is

$$|\tilde{e} - e| = O\big(\Delta x^{\min(p+q,r)}\big)\,. \tag{7}$$

The ratio

$$\frac{||\tilde{e}||}{||\tilde{u} - u||} = 1 + O\big(\Delta x^{\min(q,r-p)}\big)$$

indicates that under the assumed conditions, the error estimate will be asymptotically correct so long as $r > p$. Therefore, the use of the same orders for primal and error equation discretizations results in asymptotically correct error estimates so long as the residual approximation is sufficiently accurate, as demonstrated in [11, 17].

The evaluation of the residual has been a primary focus of investigation in the literature to date. This issue arises for FDM and FVM schemes because of the discrete nature of the approximate solution to the primal equation. As identified by Hay and Visonneau [17], there are three approaches commonly taken:

1. Approximate the residual by a discretization;
2. Approximate the residual by one or more leading truncation error terms from the modified differential equation of the primal discretization [11, 12, 14, 18];
3. Approximate the residual by reconstructing the approximate solution and directly applying the differential residual operator [13, 17, 20, 21].

The first is the standard approach for more traditional uses of defect correction [10, 20, 21] and requires the use of a different and typically high-order discretization than used in the primal equation. However, because the residual represents a local source of error as a function of the approximate primal solution and not the error itself, stability of the error discretization is largely independent of the specifics of the discrete residual operator. This discrete approximation of the residual operator is the approach we adopt because of its relative simplicity for logically rectangular Cartesian or curvilinear grids. As it relates to defect correction, the technique is not new, but we have not seen this approach employed in the error transport literature, even though it has some advantages.

In contrast, the second choice relies on the asymptotic equivalence of the residual (differential operator applied to the discrete solution) and truncation error (difference operator applied to the continuous solution). Typically, only the leading-order truncation error term is used. This approach can be challenging because the truncation error for many modern nonlinear discretizations can be difficult to derive and/or evaluate numerically. As a further complication, there are cases where the error behavior is not well represented by the first-order term and so two or more terms of the truncation error may be required. The final approach requires interpolants of order sufficient to provide both the necessary derivatives and accuracy as well as consistency with any initial or boundary conditions. We note that asymptotically, all of these approaches are equivalent, and that, in the case of linear operators, every choice of linear reconstruction (Option 3) leads to a particular choice of linear discretization (Option 1).

Before proceeding to the specifics of our residual evaluation strategy for method-of-lines and space-time discretizations, we will remark that another error transport approach exists in the literature. Our formalism is based on the continuous error equation that has been referred to in the literature as "Continuous Error Transport Equation" or "Exact Operator Residual" methods. In contrast and in a completely analogous way, one can develop a discrete error equation directly by beginning with the difference equation solved exactly by the discrete solution and introducing a discrete form of (2). In this case, the error equation is driven by the truncation error instead of the residual, but the evaluation of the source term is still an important issue [12, 14–16, 19]. Such methods have been referred to as "Discrete Error Transport Equation" or "Approximate Operator Residual" methods. Asymptotically, the two approaches are equivalent.

3. Residual Evaluation

Our approach defines an approximation of the residual that can be evaluated to *arbitrary* order without deriving any terms in the modified differential equation: we are forming a difference approximation to the continuous residual operator that is applied to the discrete approximate solution. As such, the residual is asymptotically equivalent to the truncation error terms from the primal operator approximation plus the truncation error terms from the residual operator approximation. In the time-dependent case, the residual operator contains both time and space derivatives, but we wish to avoid explicit interpolation in time that would be necessary in a direct reconstruction method [13, 17, 20, 21]. Furthermore, approximation of the residual by leading-order truncation error terms [11, 12, 14, 18] is not only scheme-specific and difficult for complex problems, but also potentially less robust for under-resolved features where neglected terms may not be negligible.

We consider the two types of temporal discretization approaches common for hyperbolic PDEs: the method-of-lines approach, where space is discretized first and then an appropriate ODE discretization is applied in time, and the space-time approach, where schemes take advantage of the coupled nature of time and space to cancel leading spatial error terms with leading temporal error terms. While the method-of-lines formulation is straightforward, the construction of a higher-order residual approximation for space-time schemes requires some additional explanation. In this section, we give a general overview of the two approaches. Specific examples of each for a model problem are provided in Section 4.

3.1. Method-of-Lines Formulation

The method-of-lines formulation provides a conceptually simple approach to the high-order discretization of time-dependent problems because of the independent treatment of space and time. Indeed, one of the original and popular approaches to the discretization of the Euler equations combined a central second-order spatial discretization with the explicit, four-stage fourth-order Runge-Kutta scheme [22]. We show that the method-of-lines formulation allows for a particularly simple evaluation of the residual.

We discretize the spatial operator of the primal equation (1) to obtain the semi-discrete form

$$\partial_t \tilde{u}_h + F_h(\tilde{u}_h) = 0, \tag{8}$$

where F_h is some discrete operator, h is the spatial discretization parameter, and $\tilde{u}_h(t)$ is a vector of time-dependent variables. Similarly, we spatially discretize the error equation:

$$\partial_t \tilde{e}_h + G_h(\tilde{e}_h; \tilde{u}_h) = -\partial_t \tilde{u}_h - F_h^*(\tilde{u}_h), \tag{9}$$

where G_h and F_h^* are discrete approximations of $\mathcal{G}$ and $\mathcal{F}$, respectively. Note that we do not assume that $G_h = F_h$. It should be made clear that the right-hand side discretization must not be the same discretization used for the primal equation, because the semi-discrete primal solution $\tilde{u}_h$ exactly solves that semi-discrete equation (8). To make this explicitly clear, using (8) in the right-hand side of (9), we find

$$\partial_t \tilde{e}_h + G_h(\tilde{e}_h; \tilde{u}_h) = F_h(\tilde{u}_h) - F_h^*(\tilde{u}_h). \tag{10}$$

The source of error vanishes for $F_h = F_h^*$. In addition, using (8) in (9) analytically eliminates the time derivative from the residual. To obtain a fully discrete scheme, one applies a suitable ODE integrator and co-evolves the primal and error approximations.

3.2. Space-Time Formulation

For hyperbolic systems, it is common to discretize the primal equation using a space-time scheme, by which we mean a one-step scheme (possibly with predictor stages) that couples the time and space discretizations. The Lax-Wendroff [23] and MUSCL-Hancock [24, 25] schemes are both of this type. The construction of space-time schemes is specific to the differential form of the spatial operator, but the general principle is that successive derivatives of the governing equation are used to exchange temporal with spatial derivatives in a Taylor series expansion in time. The resulting spatial derivatives are then discretized. A similar procedure can be used in the construction of a difference approximation to the residual operator.

Because the extension is not as clear for space-time formulations as it was for the method-of-lines in Section 3.1, we consider the more general hyperbolic system

$$\partial_t u + \partial_x f(u) = 0, \tag{11}$$

where $f(u)$ is a flux function whose Jacobian $\partial_u f = A(u)$ has real eigenvalues. Note that the approach is by no means restricted to this form of operator.

Consider the residual for a continuous approximation to the solution of (11):

$$\mathcal{R}(\tilde{u}) = \partial_t \tilde{u} + \partial_x f(\tilde{u}). \tag{12}$$

Using the approximate solution at two time levels, $\tilde{u}(x, t^n)$ and $\tilde{u}(x, t^{n+1})$, we can construct the following second-order approximation of $\mathcal{R}(\tilde{u})$ at time

$t = t^{n+1/2}$:

$$\begin{aligned}\mathcal{R}(\tilde{u}(x,t^{n+1/2})) =& \frac{\tilde{u}(x,t^{n+1}) - \tilde{u}(x,t^n)}{\Delta t} \\ &+ \partial_x \left(\frac{f\left(\tilde{u}(x,t^{n+1})\right) + f\left(\tilde{u}(x,t^n)\right)}{2} \right) \\ &+ \frac{\Delta t^2}{12} \partial_{ttt}\tilde{u}\bigg|_{t^{n+1/2}} + O\big(\Delta t^4\big)\,.\end{aligned} \tag{13}$$

Following the standard space-time procedure, we use the primal equation to convert $\partial_{ttt}u$ into spatial derivatives; specifically,

$$\partial_t u = -\partial_x f(u), \tag{14a}$$

$$\partial_{tt} u = -\partial_{tx} f(u) = -\partial_x \left[A(u)\partial_t u\right] = \partial_x \left[A(u)\partial_x f(u)\right], \tag{14b}$$

$$\begin{aligned}\partial_{ttt} u &= \partial_{tx}\left[A(u)\partial_x f(u)\right], \\ &= -\partial_x \left[\frac{\partial A}{\partial u} \left(\partial_x f(u)\right)^2 + A(u)\partial_x \left[A(u)\partial_x f(u)\right] \right].\end{aligned} \tag{14c}$$

We thus can construct a formally fourth-order in time difference approximation of the residual at $t = t^{n+1/2}$ by combining (13) and (14c):

$$\begin{aligned}\mathcal{R}(\tilde{u}(x,t^{n+1/2})) \approx& \frac{\tilde{u}(x,t^{n+1}) - \tilde{u}(x,t^n)}{\Delta t} \\ &+ \partial_x \left(\frac{f\left(\tilde{u}(x,t^{n+1})\right) + f\left(\tilde{u}(x,t^n)\right)}{2} \right) \\ &- \frac{\Delta t^2}{12} \partial_x \left[\frac{\partial A}{\partial u} \left(\partial_x f(\tilde{u})\right)^2 + A(\tilde{u})\partial_x \left[A(\tilde{u})\partial_x f(\tilde{u})\right] \right]_{t^{n+1/2}}.\end{aligned} \tag{15}$$

Applying this approach to higher temporal derivatives in the Taylor series expansion leads to higher-order residual approximations in time. Appropriate spatial differences are subsequently used to approximate the spatial derivatives.

4. Model Nonlinear Problem: Inviscid Burgers' Equation

The linear error transport approach has been applied to the steady-state [13, 17–19] and time-dependent [11] Euler and Navier-Stokes systems that describe fluid flow. To describe the nonlinear error transport approach, we first restrict our consideration to a scalar equation for clarity. The inviscid Burgers' equation,

$$\partial_t u + \partial_x \left(\frac{1}{2} u^2 \right) = 0, \qquad x \in \mathbb{R}, \quad t > 0, \tag{16}$$

is a canonical model to consider because it possesses an advective nonlinearity similar to that of the Euler and Navier-Stokes equations. We assume an initial value problem (IVP) with initial conditions given as $u(x, t = 0) = g(x)$. Solutions can be constructed using the method of characteristics; see, for instance Whitham [26].

In practice, strong solutions to (16) exist only for finite time. In particular, discontinuities can form when the solution steepens through nonlinear interaction. It is therefore common to consider solutions in a weak sense. For this reason, the divergence, or conservation, form (16) is commonly used. The continuous nonlinear error transport equation for an approximate solution $\tilde{u}(x, t)$ to the Burgers' equations is

$$\partial_t e + \partial_x \left(\frac{1}{2} e^2 \right) + \partial_x \left(\tilde{u} e \right) = -\partial_t \tilde{u} - \partial_x \left(\frac{1}{2} \tilde{u}^2 \right). \tag{17}$$

The left-hand side is a nonlinear and variable-coefficient advection operator applied to e. Both of the spatial derivative terms on the left-hand side can be large, with either one of them dominating in a particular region of the solution. As a result, both terms must be present in order to describe the time evolution of the error for some cases. The inclusion of the nonlinear term $\partial_x(e^2/2)$ distinguishes our nonlinear error transport approach from the linear error transport approach that has been used previously. In Section 6, we will demonstrate advantages of preserving the nonlinearity.

We consider conservative, upwind, finite-difference schemes [24] that are well suited to this class of problems. Both method-of-lines and space-time discretizations will be developed. Similar space-time schemes were employed in [11].

4.1. Method-Of-Lines Discretization

We explicitly specify our semi-discrete scheme for both the primal and error equations. Temporal discretization for all MOL discretizations throughout this paper is accomplished using the standard explicit, four-stage, fourth-order Runge-Kutta scheme.

4.1.1. Discretization of the Primal Equation

Consider a semi-discrete approximation of (16) on the uniform spatial grid $x_i = x_0 + i\Delta x$. Spatial approximation is indicated using a subscript (*e.g.*, $\tilde{u}_i(t) \approx \tilde{u}(x_i, t)$). A conservative spatial approximation for Burgers' equation can be written as

$$\partial_t \tilde{u}_i = -\frac{1}{2\Delta x}\Delta_+ \left[(\tilde{u}_{i-1/2})^2\right], \tag{18}$$

where $\Delta_+[v_i] = v_{i+1} - v_i$ is the forward difference operator and where the $\tilde{u}_{i\pm 1/2}$ are determined through the solution of Riemann problems [26, 27] at cell faces. Specifically, if we define $\tilde{u}_{i-1,+}$ and $\tilde{u}_{i,-}$ to be the values at the left- and right- of the $(i-1/2)$ interface, respectively,

$$\tilde{u}_{i-1/2} = \begin{cases} \tilde{u}_{i-1,+}, & \text{if} \quad \tilde{u}_{i,-} > \tilde{u}_{i-1,+} > 0, \\ \tilde{u}_{i,-}, & \text{if} \quad 0 > \tilde{u}_{i,-} > \tilde{u}_{i-1,+}, \\ \tilde{u}_{i-1,+}, & \text{if} \quad \frac{1}{2}(\tilde{u}_{i-1,+} + \tilde{u}_{i,-}) > 0 \;\text{ and }\; \tilde{u}_{i,-} \le \tilde{u}_{i-1,+}, \\ \tilde{u}_{i,-}, & \text{if} \quad \frac{1}{2}(\tilde{u}_{i-1,+} + \tilde{u}_{i,-}) \le 0 \;\text{ and }\; \tilde{u}_{i,-} \le \tilde{u}_{i-1,+}, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

To increase the spatial order of accuracy, piecewise linear reconstruction over each cell is used to obtain improved approximations to the solution values $\tilde{u}_{i,\pm}$ at cell boundaries:

$$\tilde{u}_{i,\pm} = \tilde{u}_i \pm \frac{1}{2}\psi\left(\Delta_+\left[\tilde{u}_i\right], \Delta_+\left[\tilde{u}_{i-1}\right]\right).$$

The limiter function, $\psi(\cdot,\cdot)$, is used to change the character of the scheme. We will consider three choices for ψ:

$$\psi_1(u,v) = 0, \tag{20a}$$

$$\psi_2(u,v) = \frac{1}{2}(u+v), \tag{20b}$$

$$\psi_{MM}(u,v) = \operatorname{minmod}(u,v), \tag{20c}$$

where

$$\operatorname{minmod}(u,v) = \begin{cases} u, & \text{if } |u| < |v| \text{ and } uv > 0, \\ v, & \text{if } |u| \ge |v| \text{ and } uv > 0, \\ 0, & \text{if } uv \le 0. \end{cases}$$

The first case will give a first-order upwind scheme, the second will give an unlimited second-order upwind scheme, and the third choice yields a total-variation-diminising (TVD), limited, high-resolution scheme [24, 28].

4.1.2. Discretization of the Error Equation

Consider now the numerical approximation of the error equation (17). As noted previously, in the finite-difference (finite-volume) framework, we are free to choose the discretization of the left-hand and the right-hand sides as desired under the constraint that the residual is evaluated with a different discrete operator from that used for the primal equation ($F_h \neq F_h^*$). There are other accuracy considerations as given by (7): the residual should be evaluated at higher-order than the primal equation and the error equation should be discretized at the same or higher order than the primal equation. Nevertheless, because the error does not appear in the right hand side of the error equation, we are free to choose high-order stencils without stability concerns.

The discretization of the differential residual

$$\mathcal{R}(\tilde{u}) = \partial_t \tilde{u} + \partial_x \left(\frac{1}{2} \tilde{u}^2 \right) \tag{21}$$

for Burgers' error equation (17) is straight-forward. We assume that $\tilde{u}_i$ is a discrete approximation of $\tilde{u}$ that satisfies (18). Thus, following the discussion in Section 3.1, the primal discretization (18) provides a discrete expression of the first term in (21).

We can choose a discrete approximation of the second term in (21) that uses the same 5-point stencil required for (18). There are two obvious candidate fourth-order approximations: $\Delta x \partial_x \left(\tilde{u}^2\right) \approx \Delta_0^{(4)} \left[\tilde{u}_i^2\right]$ and $\Delta x \partial_x \left(\tilde{u}^2\right) \approx 2\tilde{u}_i \Delta_0^{(4)} \left[\tilde{u}_i\right]$. Here $\Delta_0^{(4)} = \Delta_0 \left(1 - \frac{1}{6}\Delta_+ \Delta_-\right)$ is the fourth-order, undivided, first-difference operator. These two choices correspond to a conservative and a quasi-linear approximation of this term and will result in error approximations with different properties. From an accuracy perspective, a truncation error analysis reveals that

$$\frac{1}{2\Delta x} \Delta_0^{(4)} \left[\tilde{u}_i^2\right] = \partial_x \left(\tilde{u}^2\right) - \frac{\Delta x^4}{30} \left(\tilde{u} \partial_x^5 \tilde{u} + 5 \partial_x \tilde{u} \partial_x^4 \tilde{u} + 10 \partial_x^2 \tilde{u} \partial_x^3 \tilde{u}\right) + \cdots$$

while

$$\frac{1}{\Delta x} \tilde{u}_i \Delta_0^{(4)} \left[\tilde{u}_i\right] = \partial_x \left(\tilde{u}^2\right) - \frac{\Delta x^4}{30} \left(\tilde{u} \partial_x^5 \tilde{u}\right) + \cdots .$$

For smooth solutions, the latter would be a better choice because of its smaller truncation error. The discretization of the right hand side is in this case

$$\mathcal{R}(\tilde{u}_i) = -\frac{1}{2\Delta x} \Delta_+ \left[(u_{i-1/2})^2\right] + \frac{1}{\Delta x} u_i \Delta_0^{(4)} \left[u_i\right] + O\big(\Delta x^4\big) . \tag{22}$$

For weak solutions, the conservative choice may be more appropriate, which would lead to

$$\mathcal{R}(\tilde{u}_i) = -\frac{1}{2\Delta x}\Delta_+\left[(u_{i-1/2})^2\right] + \frac{1}{2\Delta x}\Delta_0^{(4)}\left[(u_i)^2\right] + O\big(\Delta x^4\big)\,. \tag{23}$$

Other valid choices could be made.

The left hand side of (17) consists of the non-linear Burgers' term $\partial_x(e^2/2)$ and the non-constant-coefficient advection term $\partial_x(\tilde{u}e)$, representing non-linear and non-constant coefficient transport of e. We note that in this nonlinear case, this is not the same operator as in the primal Burgers' equation, but it is still amenable to a similar discretization strategy.

As was done for the primal equation, we consider the full left hand side of the error equation (17) in conservative form. Using the residual discretization (22), we form the semi-discrete, conservative, upwind scheme

$$\partial_t \tilde{e}_i = -\frac{1}{\Delta x}\Delta_+\left[\tilde{e}_{i-1/2}(\tilde{u}_{i-1/2} + \frac{1}{2}\tilde{e}_{i-1/2})\right] + \mathcal{R}(\tilde{u}_i).$$

The definition of $\tilde{u}_{i\pm1/2}$ follows from the polynomial interpolant of $\tilde{u}$ from $\tilde{u}_i$ that is consistent with the discretization chosen in (22)

$$\tilde{u}_{i-1/2} = \frac{1}{16}\left(-\tilde{u}_{i-2} + 9\tilde{u}_{i-1} + 9\tilde{u}_i - \tilde{u}_{i+1}\right) + O\big(\Delta x^4\big)\,.$$

The definition of $\tilde{e}_{i\pm1/2}$ is similar to the definition of $\tilde{u}_{i\pm1/2}$ in Section 4.1.1. Approximations to the error at the right and left cell boundaries are defined as

$$\tilde{e}_{i,\pm} = \tilde{e}_i \pm \frac{1}{2}\psi\left(\Delta_+\left[\tilde{e}_i\right], \Delta_+\left[\tilde{e}_{i-1}\right]\right).$$

Definitions of $\tilde{e}_{i\pm1/2}$ follow from the solution to the Riemann problem

$$\tilde{e}_{i-1/2} = \begin{cases} \tilde{e}_{i-1,+}, & \text{if} \quad \tilde{e}_{i,-} > \tilde{e}_{i-1,+} > \tilde{u}_{i-1/2}, \\ \tilde{e}_{i,-}, & \text{if} \quad \tilde{u}_{i-1/2} > \tilde{e}_{i,-} > \tilde{e}_{i-1,+}, \\ \tilde{e}_{i-1,+}, & \text{if} \quad \frac{1}{2}(\tilde{e}_{i-1,+} + \tilde{e}_{i,-}) > \tilde{u}_{i-1/2} \ \text{ and } \ \tilde{e}_{i,-} \le \tilde{e}_{i-1,+}, \\ \tilde{e}_{i,-}, & \text{if} \quad \frac{1}{2}\tilde{e}_{i-1,+} + \tilde{e}_{i,-}) \le \tilde{u}_{i-1/2} \ \text{ and } \ \tilde{e}_{i,-} \le \tilde{e}_{i-1,+}, \\ 0, & \text{otherwise.} \end{cases}$$

4.2. Space-Time Formulation

We will use a MUSCL-Hancock approach that makes use of a spatial discretization having many features in common with our method-of-lines formulation.

4.2.1. Discretization of the Primal Equation

For Burgers' equation, we write the fully discrete, conservative updates in the form

$$\tilde{u}_i^{n+1} = \tilde{u}_i^n - \frac{\Delta t}{2\Delta x}\Delta_+ \left[\left(\tilde{u}_{i-1/2}^{n+1/2} \right)^2 \right]. \tag{24}$$

Here $\tilde{u}_{i-1/2}^{n+1/2}$ is found as the solution to the Riemann problem (19) with left and right inputs evaluated at the mid-time level. These left and right states are defined through Taylor expansion in both space and time; formally, to second-order in space and time,

$$\tilde{u}_{i,\pm}^{n+1/2} = \tilde{u}_i^n + \frac{1}{2}\left(\pm 1 - \frac{\tilde{u}_i^n \Delta t}{\Delta x} \right) \psi\left(\Delta_+ \left[\tilde{u}_i^n \right], \Delta_+ \left[\tilde{u}_{i-1}^n \right] \right).$$

4.2.2. Discretization of the Error Equation

In order for the eventual discretization of the error equation to be higher-order accurate for smooth flows (as was the case for the method of lines discretization), the discretization of the transport terms must be at least as accurate as the primal discretization. Here we are considering second-order primal discretizations, and so a second-order Hancock predictor-corrector scheme for the error equation can be written as

$$\tilde{e}_i^{n+1} = \tilde{e}_i^n - \frac{\Delta t}{2\Delta x}\Delta_+ \left[\left(\tilde{e}_{i-1/2}^{n+1/2} \right)^2 \right] - \Delta t R_i^{n+1/2}.$$

We must still specify $\tilde{e}_{i-1/2}^{n+1/2}$ and $R_i^{n+1/2} = R(\tilde{u}_i^{n+1/2})$ with sufficient accuracy.

Consider $R_i^{n+1/2}$ first. Following the approach from Section 3.2, we first construct a discretization using only two time levels. In order to provide for up to fourth order accuracy for the error, by (7), a third order accurate approximation for $\mathcal{R}(\tilde{u})$ must be devised. The primal continuous equation can be used to determine the following:

$$\partial_t^3 u = -6u(\partial_x u)^3 - 9u^2 \partial_x u \partial_x^2 u - u^3 \partial_x^3 u. \tag{25}$$

Thus, the continuous residual is discretized to third order accuracy as

$$\begin{aligned} R_i^{n+1/2} = & \frac{\tilde{u}_i^{n+1} - \tilde{u}_i^n}{\Delta t} \\ & + \frac{1}{24\Delta x}\left(\tilde{u}_i^n \Delta_0^{(4)}\left[\tilde{u}_i^n \right] + \tilde{u}_i^{n+1} \Delta_0^{(4)}\left[\tilde{u}_i^{n+1} \right] \right) \\ & + \frac{\Delta t^2}{24}\left(D_{ttt}[\tilde{u}_i^n] + D_{ttt}[\tilde{u}_i^{n+1}] \right), \end{aligned} \tag{26}$$

where

$$D_{ttt}[\tilde{u}_i^n] = -\frac{1}{\Delta x^3}\left(6\tilde{u}_i^n(\Delta_0^{(4)}[\tilde{u}_i^n])^3 + 9(\tilde{u}_i^n)^2\Delta_0^{(4)}[\tilde{u}_i^n]\Delta_{xx}[\tilde{u}_i^n] + (\tilde{u}_i^n)^3\Delta_{xxx}[\tilde{u}_i^n]\right),$$

$\Delta_{xx} = \Delta_+\Delta_-(1 - \Delta_+\Delta_-/12)$, and $\Delta_{xxx} = \Delta_0\Delta_+\Delta_-$.

Next we show how to determine the predictor state, $\tilde{e}_{i-1/2}^{n+1/2}$. A third-order estimate of the residual at time level $t = t^n$ is required for the predictor. A Taylor series expansion about $t = t^n$ allows us to define

$$R_i^n = -\frac{\tilde{u}_i^{n+1} - \tilde{u}_i^n}{\Delta t} + \frac{\Delta t}{2}D_{tt}[\tilde{u}_i^n] + \frac{\Delta t^2}{6}D_{ttt}[\tilde{u}_i^n],$$

where $D_{ttt}[\tilde{u}_i^n]$ is as before and

$$D_{tt}[u_i^n] = \frac{1}{\Delta x^2}\left(2u_i^n(\Delta_0^{(4)}[u_i^n])^2 + (u_i^n)^2\Delta_{xx}[u_i^n]\right).$$

From here it is straightforward to define left and right states for the Riemann problem using

$$\tilde{e}_{i,\pm}^{n+1/2} = \tilde{e}_i^n + \frac{1}{2}\left(\pm 1 - \frac{(\tilde{e}_i^n + \tilde{u}_i^n)\Delta t}{\Delta x}\right)\psi\left(\Delta_+\left[\tilde{e}_i^n\right], \Delta_+\left[\tilde{e}_{i-1}^n\right]\right) + \frac{\Delta t}{2}\left(R_i^n - \frac{\tilde{e}_i^n}{\Delta x}\Delta_0[\tilde{u}_i^n]\right).$$

In addition, the primal approximation at the cell face is defined through Taylor expansion, for example as

$$\tilde{u}_{i-1/2}^{n+1/2} = \frac{\tilde{u}_i^n + \tilde{u}_{i-1}^n}{2}\left(1 - \frac{\Delta t}{\Delta x}(\tilde{u}_i^n - \tilde{u}_{i-1}^n)\right).$$

Finally, the Riemann problem at cell interfaces is solved as was done for the method-of-lines case (4.1.2) with all left and right values evaluated at the mid-time level.

5. Convergence Properties

We demonstrate convergence properties of the nonlinear error transport technique with direct discretization of the residual when applied to smooth solutions for several discretizations. We verify that the method produces the predicted convergence rates for smooth solutions and linear schemes and that convergence degrades, as expected, for nonlinear schemes even on smooth problems.

5.1. Inviscid Burgers Test Problem

Consider the sinusoidal initial value problem (IVP) for Burgers equation (16):

$$u(x, t = 0) = a - \sin(\pi x). \tag{27}$$

For times $t < t_b = \pi^{-1}$, the solution remains smooth and is implicitly given by

$$u(\xi, t) = a - \sin(\pi\xi)$$

along characteristics

$$x(\xi, t) = \xi + [a - \sin(\pi\xi)]t.$$

Regions of positive slope expand while regions of negative slope compress.

At $t = t_b$, shocks form at $x_s(t) = at + 2k$ for $k \in \mathbb{Z}$. For times $t \geq t_b$, the feet of the characteristics arriving at the shock at time t are $\xi_s(t)$, which are found by solving the transcendental equation

$$\xi_s(t) = t\sin(\pi\xi_s(t)).$$

The solution to Burgers' equation is then

$$u(x, t) = \begin{cases} a - \sin(\pi\xi), & |\xi| \geq |\xi_s(t)|, \\ a - \sin(\pi\xi_s(t)), & \text{otherwise}, \end{cases} \tag{28}$$

at locations corresponding to

$$x(\xi, t) = \begin{cases} \xi + [a - \sin(\pi\xi)]t, & |\xi| \geq |\xi_s(t)|, \\ x_s(t), & \text{otherwise}. \end{cases} \tag{29}$$

The shocks grow in magnitude until $t = 1/2$, when the maxima of the solution catch the shocks, and decay thereafter.

5.2. Properties for Solution-Independent Discretizations

In Section 2, we derived an upper bound for the convergence rate of the error given the orders of the primal operator, error operator, and residual operator discretizations. To satisfy the conditions that lead to that estimate, we consider the smooth solution (28) for $t < t_b$. To avoid nonlinear upwinding, we take $a = 2$ and consider the solution at time $t = 0.1$. In

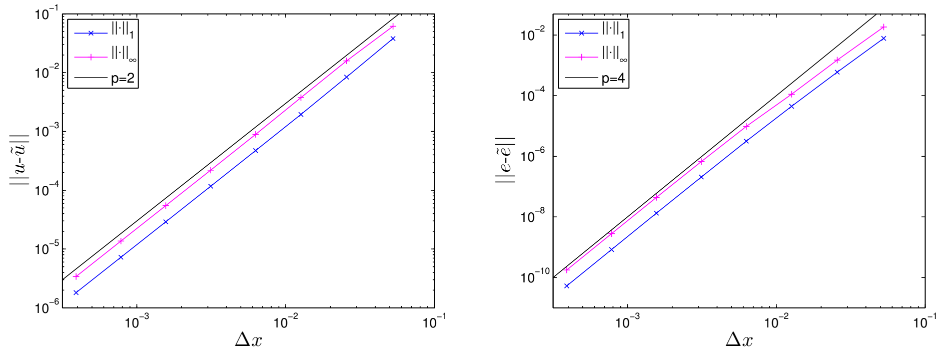

Figure 1: Convergence of the approximate solution and error at $t = 0.1$ for $u(x, t = 0) = 2 - \sin(\pi x)$ for the method-of-lines discretization of Burgers' equation.

addition, we take ψ to be defined by (20b) in order to obtain a linear, formally second-order spatial discretization. The order relation (7) predicts that $|\tilde{e} - e| = O(\Delta x^4)$ for method-of-lines and space-time discretizations, since $p = q = 2$ and either residual evaluation (22) or (23), which both have r=4.

As a computational example, we simulate on the domain $x \in [-1, 1)$ using N points, apply periodic boundary conditions, and use a time step determined by a fixed CFL restriction of 0.9:

$$\max_{i\in[1,N]} (|\tilde{u}_i^n| + |\tilde{e}_i^n|) \frac{\Delta t}{\Delta x} \leq 0.9.$$

We consider a sequence of meshes $N \in \{10 \cdot 2^m; 2 \leq m \leq 9\}$ and perform convergence studies for both the method-of-lines and space-time schemes.

Figures 1 and 2 show the results of convergence studies for the L_1 and L_∞ norms of the error in the approximate solution and the error in the approximate error at $t = 0.1$ for the method-of-lines and space-time schemes, respectively. The non-conservative form for the residual (22) was used. Second-order convergence is achieved for both norms of the approximate solution, and fourth-order convergence is achieved for both norms of the approximate error. This is a demonstration of the convergence rates as predicted by the order relation (7).

5.3. Properties for Stencil-Changing Discretizations

For upwind discretizations of solutions that change signs, the primal discretization will switch abruptly at $u = 0$, which occurs at $x = k$ for $k \in \mathbb{Z}$.

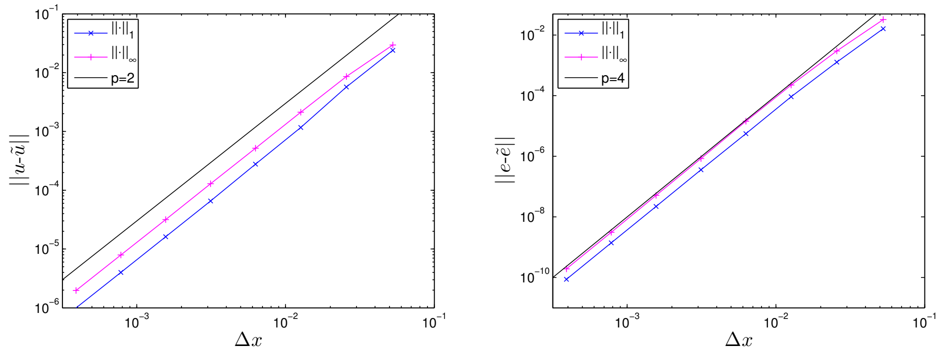

Figure 2: Convergence of the approximate solution and error at $t = 0.1$ for $u(x, t = 0) = 2 - \sin(\pi x)$ for the space-time discretization of Burgers' equation.

As shown in Figure 3, the result is that the error develops a kink or discontinuity at these points, depending on the sign of $\partial_x u$. Thus, the order relation (7) is no longer strictly applicable.

Figure 4 shows convergence study results for this case for the method-of-lines discretization using the non-conservative residual approximation (22). The primal approximate solution converges as expected at a second-order rate in both the L_1 and L_∞ error norms. However, while the L_1 norm of the error still attains fourth-order convergence, the L_∞ error norm convergence is reduced to a third-order rate. Similar results are obtained for the space-time discretization. Degraded convergence is expected in the presence of singularities [29]. We note that the error estimate is still asymptotically correct in any L_p-norm.

5.4. Properties for Limited Discretizations

From a practical perspective, one of the primary cases of interest is when nonlinear slope or flux limiting algorithms are used. We investigate the case where the reconstruction function is nonlinearly dependent on its arguments: ψ defined by (20c). The use of the limiter can result in approximate solutions with degraded accuracy even when the exact solution is perfectly smooth. Numerical errors that are not infinitely differentiable result, which has practical implications for the accuracy of the error evolution technique.

Take the initial condition (27) with $a = 1$. We consider the method-of-lines formulation with the non-conservative residual approximation; results

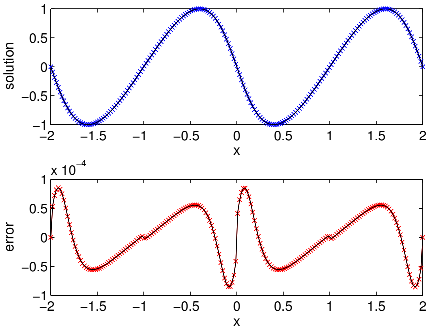

Figure 3: Plots of the approximate solution and error at $t = 0.1$ and $N = 200$ points for initial data $u(x, t = 0) = -\sin(\pi x)$ for the method-of-lines discretization of Burgers' equation. The exact solution and exact error are indicated by black lines, while the marks indicate the approximate results.

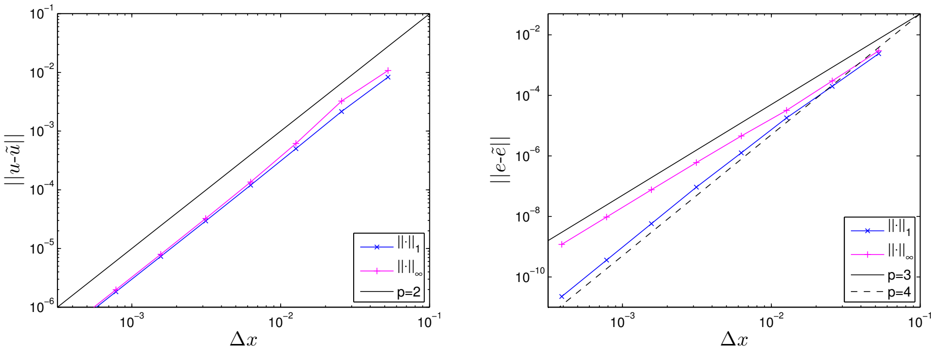

Figure 4: Convergence of the approximate solution and error at $t = 0.1$ for $u(x, t = 0) = -\sin(\pi x)$ for the method-of-lines discretization of Burgers' equation.

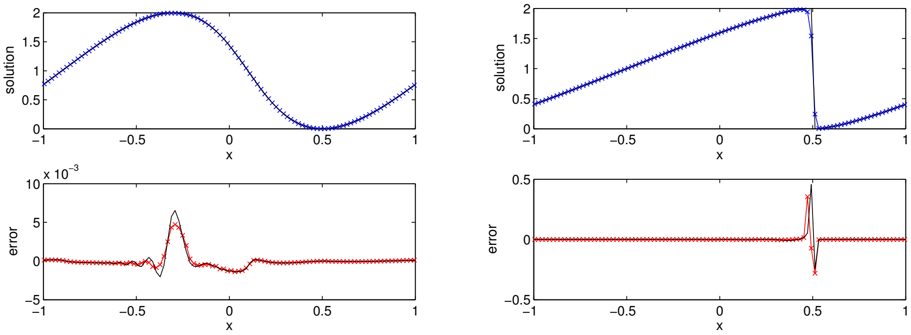

Figure 5: Solution and error of the Burgers' equation for initial data $u(x, t = 0) = 1 - \sin(\pi x)$ at $t = 0.1$ (left) and $t = 0.5$ (right). The method-of-lines scheme with the MinMod limiter was used for both the primary and error evolution equations on a grid of $N = 100$ points in $x \in [-1, 1)$. The non-conservative residual approximation was used. Black lines indicate exact values while marks indicate discrete approximations.

for the space-time scheme are similar. In Figure 5, the discrete approximation and approximate error using $N = 100$ points for $x \in [-1, 1)$ are shown at a time when the exact solution is smooth ($t = 0.1$) as well as after a shock has formed ($t = 0.5$). Prior to shock formation, the main feature in the error is associated with extrema clipping at the maximum value; a similar error does not occur at the minimum because the solution value is zero at that point. The error approximation captures this feature well. After shock formation, the error is dominated by $O(1)$ error near the captured numerical shock. The approximate numerical error represents this feature well, albeit with an $O(h)$ error in the location.

Figure 6 shows the results of a convergence study for the approximate solution. Prior to shock formation, the large error introduced near the extrema impacts the L_∞ error norm most, and we see convergence rates that trend towards a value of $4/3$.[1] The L_1 error norm is less sensitive and asymptotically tends to the expected rate of two. Other p-norms with $1 < p < \infty$ will exhibit convergence rates between these two extremes. After the formation of a shock in the flow, the L_∞ error norm is no longer convergent, which is

[1] Interestingly, this is the expected rate of convergence near a slope discontinuity for a second order scheme [29].

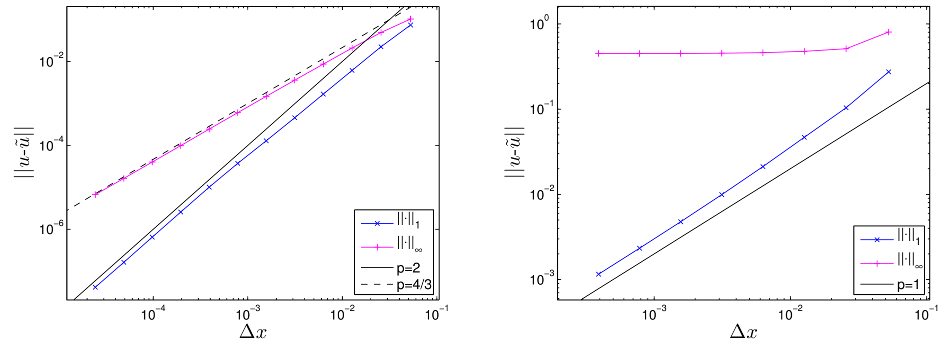

Figure 6: Approximate solution convergence for Burgers equation with initial data $u(x, t = 0) = 1 - \sin(\pi x)$ at $t = 0.1$, before shock formation (left), and at $t = 0.5$, after shock formation (right). The method-of-lines discretization and the MinMod limiter were used for both the primal and error equations. The non-conservative residual approximation was used.

to be expected because of the point-wise $O(1)$ error in the neighborhood of the captured shock. On the other hand, the L_1 error norm converges at the expected first-order rate.

The asymptotic behavior of the approximate error is shown in Figure 7. For smooth flows, the L_∞ error norm of the error converges near the $4/3$ rate that was seen for the approximate solution, but very high resolution ($N > 20000$) is needed to achieve this. Likewise, L_1 convergence rate of the approximate error approaches a second-order rate, but only for very high spatial resolutions. After shock formation, the L_1 convergence of the approximate error exhibits first-order convergence while the L_∞ error norm does not converge.

Since the errors in the approximate solution and approximate error are converging at the same rates, the error estimate cannot be asymptotically correct in any L_p norm. Nevertheless, the fidelity of the approximate solutions suggests that the estimated error can still be useful. For instance, p-norm convergence of the error estimate on sub-domains away from the shock can be asymptotically correct. Preliminary work also indicates that error estimates for many other QoI constructed from the error field are asymptotically correct, depending on the sensitivity to the regions of large error.

Finally, we note that these results are robust for any choice of the offset a in the initial condition. Thus, the role of the nonlinear limiters is more

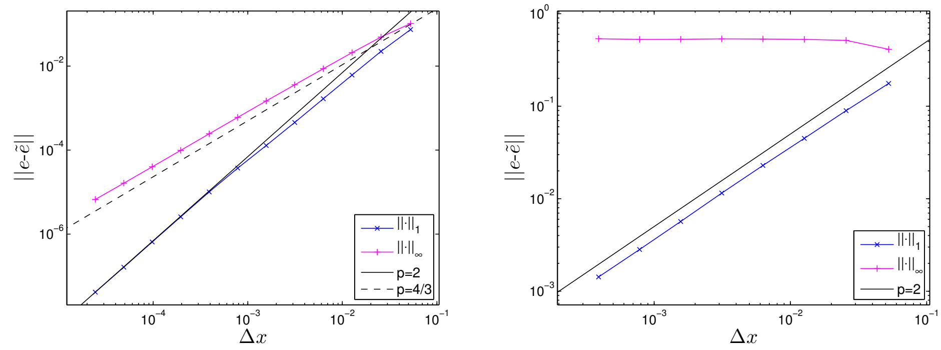

Figure 7: Approximate error convergence for Burgers equation with initial data $u(x, t = 0) = 1 - \sin(\pi x)$ at $t = 0.1$, before shock formation (left), and at $t = 0.5$, after shock formation (right). The method-of-lines discretization and the MinMod limiter were used for both the primal and error equations. The non-conservative residual approximation was used.

important than the role of nonlinearities from upwind switching, due in part to the fact that the solution is near zero when the upwind direction switches. In contrast, the solution can be any size near extrema that cause algorithm switches.

6. The Effects of Error Nonlinearity

In the existing error transport literature, the error equation operator is always linearized. For Burgers' equation, the corresponding linear error transport equation is

$$\partial_t e + \partial_x (\tilde{u} e) \approx -\partial_t \tilde{u} - \partial_x \left(\frac{1}{2} \tilde{u}^2 \right) . \tag{30}$$

The justification for applying (30) is that the $e\partial_x e$ term is higher-order and is thus negligible. Hay and Visonneau [17] investigate eliminating the $e\partial_x \tilde{u}$ term as well and determine that they do not obtain an asymptotically correct error estimate unless they use the form (30). However, even this form is not consistent with the original nonlinear error equation and cannot produce asymptotically correct results if e is large. Errors may not be small if the primal scheme is of low-order, if singularities exist in the solution, if very long-time integrations are sought, or if solution features are under-resolved. If one can handle numerically the nonlinearity of the primal equation, one should be able to handle the nonlinearity in the error equation. Our conjecture is that solving the fully nonlinear error equation will make the error estimates more robust. We demonstrate differences between linear and nonlinear error transport using the inviscid Burgers' equation.

6.1. Low-Order Approximations

One case where ignoring error nonlinearity produces unsatisfactory results is when a low-order forward approximation is used but the approximate error is sought to higher accuracy. This, for example, was done in [11]. We define the primal scheme using the ψ defined by (20a) and the error transport scheme using the ψ defined by (20b). Thus, the forward approximation is the first-order upwind method, while the error approximation uses the second-order upwind method. A conservative numerical approximation to (30) is easily defined as before, and the solution to the Riemann problem requires only the sign of the face values $\tilde{u}_{i\pm1/2}$. The method-of-lines formulation with non-conservative residual approximation is used for both equations.

Under the assumptions used to derive the order relation (7), neglecting the $e\partial_x e$ term is akin to making an additional $O(e^2) = O(\Delta x^{2p})$ error. Thus, the order relation for the linear error transport is

$$|\tilde{e} - e| = O\big(\Delta x^{\min(2p,p+q,r)}\big) . \qquad (31)$$

For $p = 1$, $q = 2$, and $r = 4$, we therefore expect second-order convergence for the linear error transport method and third-order convergence for the nonlinear error transport method for the Burgers equation. For other nonlinear equations, neglecting nonlinear error terms could pose even more serious restrictions.

Results of a convergence test for the initial condition (27) with $a = 2$ at time $t = 0.1$ are presented in Figure 8. The effect of neglecting the nonlinear term is to reduce the order of the error approximation from third to second order in both L_1 and L_∞ norms. By leaving out the $e\partial_x e$, one is limited to second order accuracy; in principle, one can obtain arbitrary accuracy with its inclusion.

6.2. Regions of Large Error

Error nonlinearity is also important when the solution develops non-differentiable features. For cases where a shock is captured, the numerical

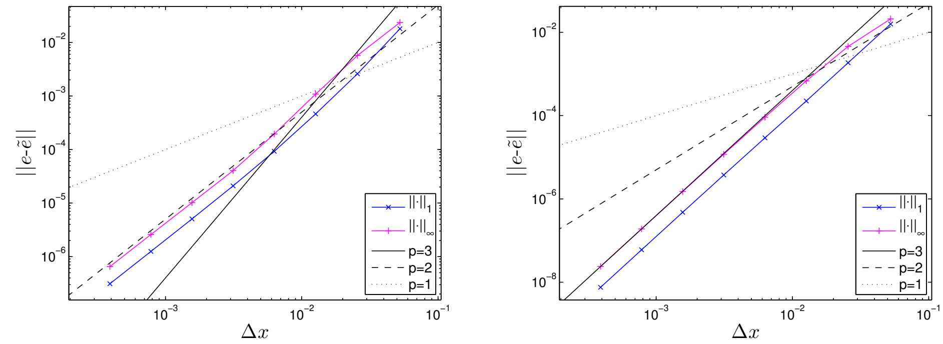

Figure 8: Convergence of the approximate solution and error at $t = 0.1$ for $u(x, t = 0) = 2 - \sin(\pi x)$ for the Burgers' equation using a method-of-lines discretization. The primal operator spatial discretization is first-order, while the error operator spatial discretization is second-order. At left are results found using the linearized equation (30), while at right are results found using the fully nonlinear error equation (17).

error is as large as the solution itself, and so nonlinear error effects are equally important as the pure linear transport of error.

6.2.1. Large Error in One Dimension

To illustrate this point in a single space dimension we again use the initial condition (27) with $a = 0$ and consider the solution at time $t = 0.5$, after a shock forms. The schemes are defined using the ψ defined by (20a), which leads to first-order upwind discretizations. The residual is evaluated in quasi-linear form as in (22). Results are presented in Figure 9.

When the linearized error transport operator of (30) is used, the error is divergent in the L_∞ error norm and not convergent in the L_1 error norm. On the other hand, the approximate error found using the nonlinear error evolution operator of (17) converges at a first-order rate in the L_1 error norm and is non-convergent in the L_∞ norm. For the linear error transport equations, the error at the stationary shock can grow with time because the residual can act as a continued source for error that the linear operator incorrectly transports. In contrast, the nonlinear transport equations provide a consistent mechanism to move error into the shock where it is destroyed by cancellation.

The potential for unbounded growth when using the linear operator in one dimension is somewhat sensitive to the choice of residual evaluation and

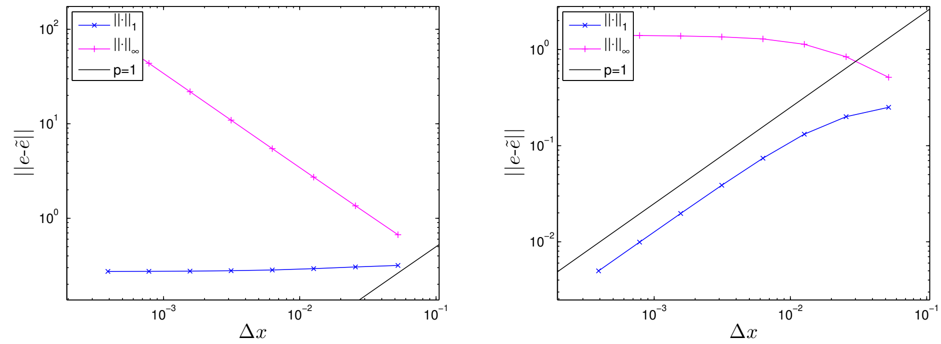

Figure 9: Convergence of the approximate error using linear error transport (left), and approximate error using nonlinear error evolution. The residual is evaluated in quasilinear form. Results are presented at $t = 0.5$ for Burgers' equation with initial data $u(x, t = 0) = -\sin(\pi x)$.

scheme selection. For example, conservative residual evaluation (23) will fix the divergent behavior of the linear error transport scheme for the first-order (ψ as in (20a)) or TVD (ψ as in (20c)) schemes. However, divergent behavior still exists for the second order upwind scheme (ψ as in (20b)).

We have run many different permutations of transport and residual formulations on many different one-dimensional problems with discontinuities. Sometimes we see growth in the error at shocks for the linear error transport method and sometimes we do not. The use of the conservative residual formulation tends to improve robustness for the linear error transport approach; a similar behavior was observed by Zhang *et al.* [11]. Nevertheless, one trend is clear: we have never seen such an unbounded growth in the error using the nonlinear error transport operator. The mere possibility of such unbounded growth should be sufficient justification for solving the fully consistent nonlinear error equation in the presence of discontinuities.

6.2.2. Large Error in Higher Dimensions

Application of the error transport approach in two space dimensions reveals further motivation for the inclusion of the nonlinear error terms. In two dimensions, the poor performance of the linear error transport method does not appear to be restricted to isolated stationary shock problems nor ameliorated by a particular choice of the residual form.

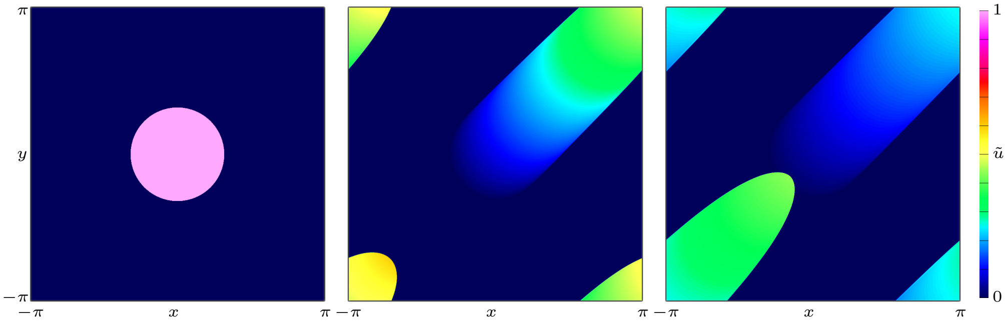

Figure 10: Approximate solution at times $t = 0$, $t = 8$, and $t = 15$ using 1601 points in the x- and y- directions.

The extension of the algorithm to a scalar, two-dimensional Burgers' equation,

$$\partial_t u + \partial_x \left(\frac{1}{2} u^2 \right) + \partial_y \left(\frac{1}{2} u^2 \right) = 0, \tag{32}$$

on the periodic domain $(x, y) \in [-\pi, \pi] \times [-\pi, \pi]$ for $t > 0$ is straightforward. We use the method-of-lines approach with the spatial discretizations (18)-(19) applied dimension-by-dimension, and the residual is evaluated in conservative form (23). Verification of the properties for smooth solutions was carried out using the method of manufactured solutions [2] and grid convergence studies as before.

We consider a discontinuous problem defined by the initial conditions

$$u(x, y, t = 0) = \begin{cases} 1 & \text{if } x^2 + y^2 < 1 \\ 0 & \text{else.} \end{cases}$$

Figure 10 shows the time evolution of this problem for $t = 0$, $t = 8$, and $t = 15$ using 1601 points in the x- and y- directions ($\Delta x = \Delta y = \pi/800$). The second-order unlimited scheme (20b) is used to avoid subtleties associated with nonlinearity in the discretization. The results are similar for the nonlinearly limited scheme.

Both the linear and nonlinear error transport techniques were applied to estimate the errors at $t = 15$. Color surface plots of the approximate error $\tilde{e}$ as well as slices along $y = -.5$ for $x \in [-2, 0]$ are provided in Figure 11. While both curved shock fronts demonstrate large error, the error estimate for linear transport demonstrates unbounded growth, as evidenced by the spike. In Figure 12, it is shown that the maximum estimated error for linear error transport grows as the mesh is refined, whereas the maximum estimated error for nonlinear error transport does not. Furthermore, we note that in this case conservative treatment of the residual did not eliminate problems with the linear transport approach.

In a single space dimension, it was only in the isolated case of stationary shocks where neglecting the error nonlinearity was found to be problematic. In two space dimensions, the behavior is richer; the error can evolve along moving or stationary shock fronts in addition to being advected into them. Since the numerical error of a shock-capturing scheme is typically $O(1)$ near a shock, nonlinear dynamics along the shock front can therefore be critical. In this problem, the nonlinear evolution along the front keeps the error estimate from growing unbounded as the grid is refined.

7. Application to the Euler Equations

We have thus far limited our application to a scalar nonlinear hyperbolic equation. In this section, we briefly consider the one-dimensional ideal gas Euler equations of fluid dynamics, which represent a hyperbolic system of nonlinear equations. While treating a system with the nonlinear error transport approach is no different *per se*, we do note that the Burgers' equation possesses a particularly amenable nonlinearity; its quadratic form produces a particularly simple form of $\mathcal{G}(\tilde{e};\tilde{u})$ in the nonlinear error equation (4).

In contrast, the Euler equations for even a simple ideal gas also contain cubic or inverse nonlinearities. Obvious approaches to evaluate $\mathcal{G}(\tilde{e};\tilde{u})$ can lead to violations of auxiliary constraints, such as the positivity of mass, density or pressure. We will show that a simple solution to this problem is to work in a flux-divergence form and to use an approximate Riemann solver based solely on the approximate primal solution to select the upwind state.

The one-dimensional Euler equations are a system of conservation laws of the form (11) with state vector $u = (\rho, \rho v, \rho E)$ and flux vector $f(u) = (\rho v, \rho v^2 + p, \rho v H)$. Here, ρ is the mass density; v is the velocity; p is the pressure; $E = e + v^2/2$ is the total specific energy; and $H = E + p/\rho$ is the total specific enthalpy. The ideal equation of state $p = (\gamma - 1)\rho e$ is used to close the system of equations.

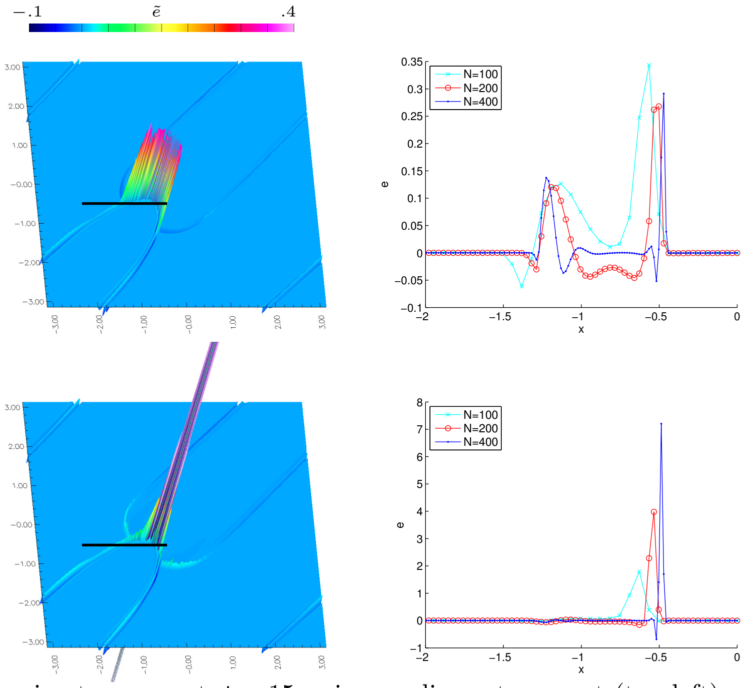

Figure 11: Approximate errors at $t = 15$ using nonlinear transport (top left), and linear transport (bottom left). The color map for the linear transport at bottom right has been restricted so that large errors are clipped out of the figure. Line plots along $y = -.5$ (indicated by the black lines on the left) are shown for each approach at right. Results are shown using 401 grid points in the x- and y- directions.

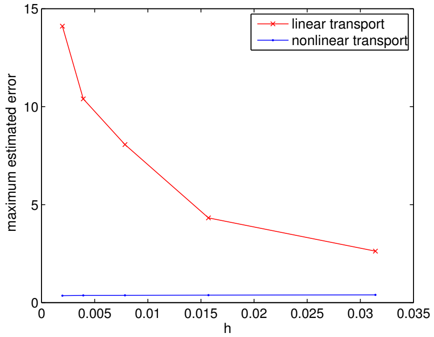

Figure 12: Maximum estimated error (in absolute value) as a function of mesh spacing. Here both linear and nonlinear transport are depicted and the divergent behavior of the linear algorithm is clear.

7.1. Discretization

For simplicity, we will use a method-of-lines approach. As in Section 4.1, we use a conservative discretization for the primal equation with an exact Riemann solver (See, *e.g.*, Toro [27]) and limited piecewise-linear reconstruction with the minmod limiter (20c). This results in a primal approximation of the form

$$\partial_t \tilde{u}_i = -\frac{1}{\Delta x} \Delta_+ \left[f(\tilde{u}_{i-1/2}) \right].$$

As previously for Burgers' equation, the semi-discrete approximation of the error equation is

$$\partial_t \tilde{e}_i = -\frac{1}{\Delta x} \Delta_+ \left[f(\tilde{u}_{i-1/2} + \tilde{e}_{i-1/2}) - f(\tilde{u}_{i-1/2}) \right] - \mathcal{R}(\tilde{u}_i) \tag{33}$$

The primary question then becomes the evaluation of the discrete flux $f(\tilde{u}_{i-1/2} + \tilde{e}_{i-1/2}) - f(\tilde{u}_{i-1/2})$. Therefore, assume a face value $\tilde{u}_{i-1/2}$ given by continuous polynomial interpolation, and left and right states $e_{i-1,+}$ and $e_{i,-}$ given by limited piecewise liner reconstruction are known. Determine the eigendecomposition $R\Lambda R^{-1} = \partial_u f(\tilde{u}_{i-1/2})$. Notice that by using the state $\tilde{u}_{i-1/2}$ rather than averages of the primal solution and face approximations of the error, the possibility of negative densities, pressures, or other unphysical states is eliminated. We now compute characteristic quantities $w_{i-1,+} = R^{-1}e_{i-1,+}$ and $w_{i,-} = R^{-1}e_{i,-}$ and compute the solution to the Riemann problem as

$$w^{(k)}_{i-1/2} = \begin{cases} w^{(k)}_{i-1,+} & \text{if} \quad \Lambda^{(k,k)} > 0 \\ w^{(k)}_{i,-} & \text{else} \end{cases}$$

for $k = 1, 2, 3$. Here the superscripts (k), or (k, k) indicate entries in the vector or matrix respectively. Finally we compute the face solution as $\tilde{e}_{i-1/2} = Rw_{i-1/2}$ which completes the description. We note that this procedure amounts to a particular choice of linearized Riemann solver, but by using (33), the full nonlinear evolution of the error is retained. Finally, we note that a conservative, fourth-order approximation of the residual is employed.

7.2. Application

For a demonstration, we consider a shock-tube IVP problem on $x \in \mathbb{R}$ with piecewise-constant initial data:

$$u = \begin{cases} u_L, & x < 1/2, \\ u_R, & x > 1/2. \end{cases} \tag{34}$$

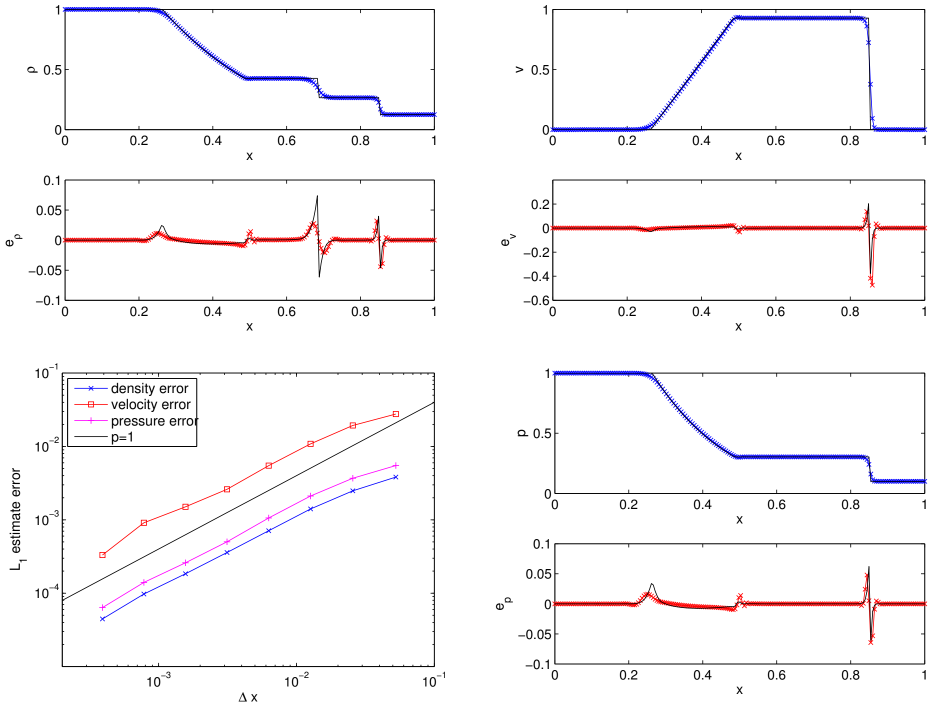

Figure 13: Approximation at time $t = 0.5$ for Sod's shock-tube problem using nonlinear error transport. Here second-order TVD algorithms have been used for spatial discretization of both primal and error equations and RK-4 for time integration. Clockwise from top left are the density, velocity, pressure, and estimated error convergence study.

We define the left and right states in primitive variables to be $(\rho_L, v_L, p_L) = (1, 0, 1)$ and $(\rho_R, v_R, p_R) = (0.125, 0, 0.1)$, which defines the standard Sod problem. The exact solution, which consists of a leftward-traveling expansion fan, a slow rightward-traveling contact discontinuity, and a faster rightward-traveling shock, can be found in [27].

We simulate on the domain $x \in [0, 1]$ until the time $t = 0.5$ so as to prevent the waves in the solution from reaching the domain boundaries. Thus, we exactly specify the boundary conditions using the piecewise constant initial data. We use the standard mesh of 200 points and plot results for a fixed CFL number of 0.4, which has no particular significance.

The results, shown in Figure 13, are quite similar to those obtained for the inviscid Burgers' equation. In this case, the error in the shock is captured very well, and the corners of the expansion fan have roughly the same degree of fidelity as seen for Burgers'. The contact discontinuity in the density field is the source of dominant error in the density field, and since this is a linearly degenerate wave with no self-steepening mechanism, capturing schemes will always smear out this wave. We see that the nonlinear error transport scheme does as well as could be expected in representing the sharp feature in the error at the contact. We also present a convergence study for the L_1 errors in the estimated errors. Near first order convergence is demonstrated for the grid resolutions presented. These results demonstrate that the nonlinear error transport approach generalizes to nonlinear systems with more complicated nonlinearities than that in Burgers' equation.

8. Conclusions

In this paper, we have investigated the nonlinear error transport method as a technique to estimate the discretization error for finite volume and finite difference discretizations of nonlinear hyperbolic equations. Our motivation is to obtain quantitative error estimates for use in uncertainty quantification. We used the scalar, time-dependent inviscid Burgers' equation as a canonical model and considered both continuous and discontinuous solutions. In contrast to prior linear transport approaches, our technique is novel because we preserved the nonlinear terms in the error equation and used a direct, systematic approach to the discretization of the error residual operator that does not require knowledge of the modified differential equation of the discretization of the primal equation. We demonstrated the approach for both method-of-lines and space-time formulations, considered linear and nonlinear discretizations as well as conservative and quasi-linear residual discretizations. Finally, the method was applied to the unsteady Euler equations in one spatial dimension, which provided a demonstration of an approach to handling more complicated nonlinearities.

Our results demonstrate that, on the whole, the method is a reasonable approach. For strong solutions, the results are in excellent agreement with the asymptotic theory (7), consistent with previous linear error transport results. Nonlinear schemes and weak solution features such as solution and slope discontinuities degrade the convergence of the error estimates. From a point-wise perspective, the computed error occurs generally in the correct location, but the magnitudes will be incorrect in the vicinity of weak solution features.

In contrast to linear transport, numerical experiments suggest that, when properly formulated, nonlinear error transport can provide bounded estimates of error, whereas linear transport schemes may allow for unbounded growth in the error near a shock. Simulations in both one and two space dimensions provide supporting evidence. In addition, the linear transport approach formally can produce a lower-order error estimate even in smooth regions of the solution. For these reasons, we advocate using the fully nonlinear error transport equation with a conservative residual discretization, particularly for problems with weak solutions.

Beyond further investigation of the limitations of nonlinear error transport for weak solutions computed with higher-order primal schemes, there are several other directions we will consider in future work. The inclusion of non-periodic boundary conditions and extension to systems in multiple dimensions should be straight-forward. Incorporating operator and dimensional splitting is another issue for investigation. In addition, coupling different primal models across domain boundaries is a logical extension of these techniques. Further developments will be sought for theoretical bounds on the amplitude errors for different weak solution features and understanding the asymptotic nature of errors in quantities of interest constructed from computed error fields.

References

[1] R. D. Skeel, Thirteen ways to estimate global error, Numer. Math. 48 (1986) 1–20.

[2] P. J. Roache, Verification and Validation in Computational Science and Engineering, Hermosa Publishers, Albuquerque, NM, 1998.

[3] C. J. Roy, Review of Discretization Error Estimators in Scientific Computing, AIAA Paper 2010-126, 2010.

[4] L. F. Richardson, The deferred approach to the limit. Part I. Single lattice, Tansactions of the Royal Society of London, Series A 226 (1927) 299–361.

[5] F. M. Hemez, J. S. Brock, J. R. Kamm, Non-linear Error Ansatz Models in Space and Time for Solution Verification, AIAA Paper 2006-1995, 2006.

[6] M. Ainsworth, J. T. Oden, A posteriori error estimation in finite element analysis, Comput. Method Appl. M. 142 (1997) 1–88.

[7] M. B. Giles, E. Süli, Adjoint methods for PDEs: *a posteriori* error analysis and postprocessing by duality, Acta Numerica 11 (2002) 145–236.

[8] D. Estep, M. Larson, R. Williams, Estimating the error of numerical solutions of systems of nonlinear reation-diffusion equations, Mem. Am. Math. Soc. 696 (2000) 1–109.

[9] I. Babuška, T. Strouboulis, C. S. Upadhyay, A model study of the quality of a posteriori error estimators for linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangles, Comput. Method Appl. M. 114 (1994) 307–378.

[10] H. J. Stetter, The defect correction principle and discretization methods, Numer. Math. 29 (1978) 425–443.

[11] X. D. Zhang, J.-Y. Trépanier, R. Camarero, A posteriori error estimation for finite-volume solutions of hyperbolic conservation laws, Comput. Method Appl. M. 185 (2000) 1–19.

[12] Y. Qin, T. I.-P. Shih, A Method for Estimating Grid-Induced Errors in Finite-Difference and Finite-Volume Methods, AIAA Paper 2003-0845, 2003.

[13] B. van Straalen, A Posteriori Error Estimation for Finite Volume Simulations of Fluid Flows, Master's thesis, University of Waterloo, 1996.

[14] Y. Qin, T. I.-P. Shih, A Discrete Transport Equation for Error Estimation in CFD, AIAA Paper 2002-0906, 2002.

[15] Y. Qin, T. I.-P. Shih, Analysis and Modeling of the Residual in Discrete-Error-Transport Equation, AIAA Paper 2003-3850, 2003.

[16] I. Celik, G. Hu, Single grid error estimation using error transport equation, J. Fluid Eng.-T. ASME 126 (2004) 778–790.

[17] A. Hay, M. Visonneau, Error estimation using the error transport equation for finite-volume methods and arbitrary meshes, Int. J. Comput. Fluid D. 20 (2006) 463–479.

[18] C. Ilinca, X. D. Zhang, J.-Y. Trépanier, R. Camarero, A comparison of three error estimation techniques for finite-volume solutions of compressible flows, Comput. Method Appl. M. 189 (2000) 1277–1294.

[19] Y. Qin, P. S. Keller, R. L. Sun, E. C. Hernandez, C.-Y. Perng, N. Trigui, Z. Han, F. Z. Shen, T. Shieh, T. I.-P. Shih, Estimating Grid-Induced Errors in CFD by Discrete-Error-Transport Equations, AIAA Paper 2004-656, 2004.

[20] N. A. Pierce, M. B. Giles, Adjoint and Defect Error Bounding and Correction for Functional Estimates, AIAA Paper 2003-3846, 2003.

[21] N. A. Pierce, M. G. Giles, Adjoint and defect error bounding and correction for functional estimates, J. Comput. Phys. 200 (2004) 769–794.

[22] A. Jameson, W. Schmidt, E. L. Turkel, Numerical Simulation of the Euler Equations by Finite Volume Methods Using Runge-Kutta Time Stepping Schemes, AIAA Paper 81-1259, 1981.

[23] P. Lax, B. Wendroff, Systems of conservation laws, Comm. Pure Appl. Math. 13 (1960) 217–237.

[24] B. van Leer, Towards the ultimate conservative difference scheme, V. A second-order sequel to Godunov's method, J. Comput. Phys. 32 (1979) 101–136.

[25] B. van Leer, On the relation between the upwind-differencing schemes of Godunov, Engquist-Osher and Roe, SIAM J. Sci. Statist. Comput. 5 (1984) 1–20.

[26] G. B. Whitham, Linear and Nonlinear Waves, Wiley-Interscience, New York, 1974.

[27] E. F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, Springer, Berlin, 1999.

[28] A. Harten, High resolution schemes for hyperbolic conservation laws, J. Comput. Phys. 49 (1983) 357–393.

[29] J. W. Banks, T. D. Aslam, W. J. Rider, On sub-linear convergence for linearly degenerate waves in capturing schemes, J. Comput. Phys. 227 (2008) 6985–7002.