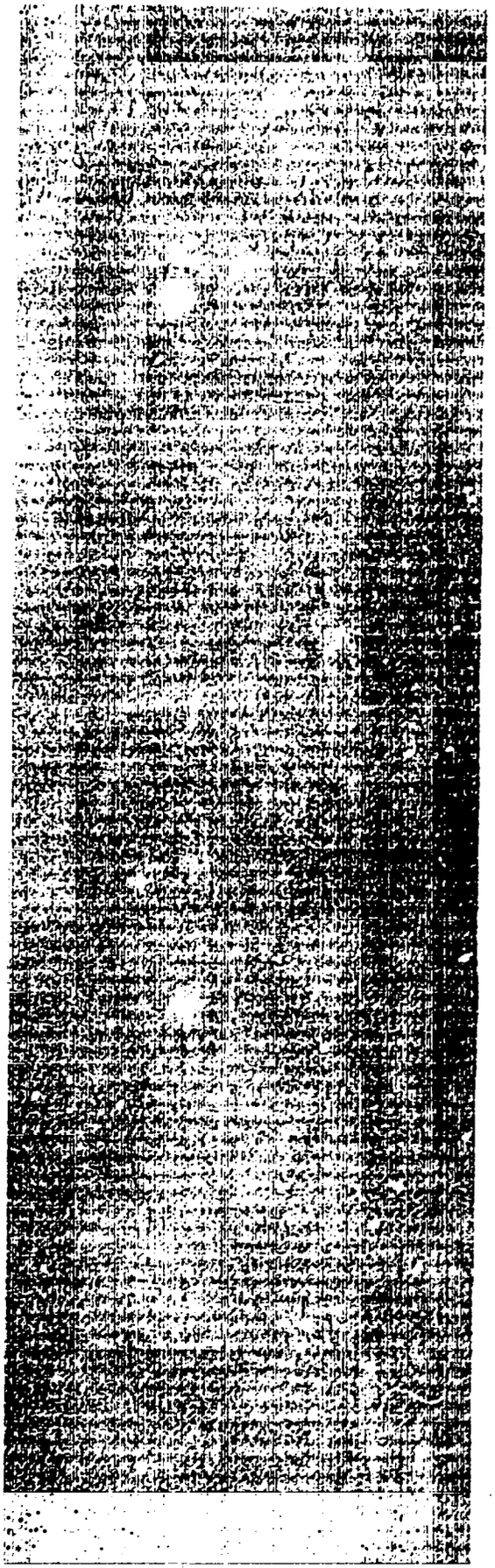
LA-12327-T
Thesis



The Design of High-Resolution Upwind Shock-Capturing Methods

Los Alamos

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

This thesis was accepted by the Department of Chemical and Nuclear Engineering, The University of New Mexico, Albuquerque, New Mexico, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. It is the independent work of the author and has not been edited by the IS-11 Writing and Editing staff.

An Affirmative Action/Equal Opportunity Employer

LA--12327-T

DE92 016866

The Design of High-Resolution Upwind Shock-Capturing Methods

William Jackson Rider

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Acknowledgements

Work such as this is never done in a vacuum. I wish to thank those who have helped, inspired, guided, entertained and otherwise assisted the completion of this research. I cannot name every person who had an impact on this work, but I will try.

While at Los Alamos National Laboratory, I have been blessed with the support of my superiors. The support of both Michael Cappiello, my section leader and and John Ireland my group leader have made the timely completion of this work possible. Others have also contributed, Drs. John Turner and Mike Hall have provided patient and excellent advice from time to time when I have managed to utterly confuse myself. They also introduced me to LaTeX, which has proved invaluable to me in the past several years. Dr. Susan Woodruff has provided some very educational advice along the course of our work together on several complex projects.

Through the years several teachers have given me something extra. Their efforts have enriched my life and shown me the value of education. My second grade teacher, Mrs. Glore, showed me the joy of learning and striving to reach one's potential. In the seventh grade one of my teachers, Mr. Preuss, showed enough concern to provide me with some extra tutoring to improve my awful spelling. My spelling is by no means wonderful, but I shutter to think what it might be without his efforts. Finally, my high school wrestling coach, Scott Evans showed me what hard work and integrity could give a person. His values proved that someone could be a winner without compromising their principles.

Some credit for this work belongs to Dr. Mohammed S. El-Genk. It was his teaching that convinced me to go to graduate school. As my master's thesis advisor, his efforts eventually pushed me to get interested in numerical methods for fluid flow.

I would like to thank the members of my committee for their patience. For reading this horrendously long dissertation and for making as many comments as they did. A number of their suggestions have made invaluable in improving the final form of this document. Dr. Norman Roderick, as chairman, was essential in helping me navigate the bureaucratic jungle that is UNM. Dr. Dennis Liles as my primary advisor gave me guidance whenever I asked for it and was always a great sounding board for my ideas.

The Design of High-Resolution Upwind Shock-Capturing Methods

by
William Jackson Rider

Abstract

The design and construction of high-resolution upwind shock-capturing methods is an effective means of solving conservation laws of physics numerically. In the past, the design of such methods was generally categorized into several distinct methods. This work shows how these methods can be viewed in a unified manner. Thus, the various types of methods can more easily take ideas from one another to improve their design.

A generalized flux-corrected transport (FCT) algorithm is shown to be total variation diminishing (TVD) under some conditions. The new algorithm has improved properties from the standpoint of use and analysis. Results show that the new FCT algorithm performs better than the older FCT algorithms and is comparable with other modern methods. This is shown to be especially important for systems of equations. The new formulation allows Riemann solvers to be used effectively with FCT methods. This directly leads to a geometric analog to symmetric TVD and FCT methods that is developed and expanded upon. This unifies these methods with high-order Godunov (HOG) methods. Two new variants of this method are derived, and shown to be uniformly non-oscillatory.

Limiters are an effective means of designing these types of methods. Earlier work by Sweby concentrated on a small set of limiters in relation to one specific difference scheme. In this research, more general classes of limiters are discussed with extensions to a wider class of schemes. In addition, flux-corrected transport and total variation bounded (TVB) limiters are discussed, modified, and expanded. Two new classes of limiters are described: s–limiters and generalized average limiters. The recently defined ULTIMATE limiter is analyzed within the framework of the other limiters. Some insight on the properties of this limiter is shown. The benefits of relaxing strict constraints on the limiters such as TVD requirements are also discussed. For coarse grids, limiters such as the TVB and the generalized average with bias improve resolution considerably. This advantage does not hold as grids are refined, because TVD-type limiters have an advantage in terms of convergence.

Lastly, the question of whether the polynomial reconstruction technique used in a HOG method should be based on cell-averages or point-values is studied. Despite

strong theoretical support of the cell-average based method, point-value reconstruction does work quite well in practice. This question is considered from two standpoints: the efficiency or economy of the reconstruction, and the accuracy and quality of the solution. The general behavior of the cell-average reconstruction is slightly more effective than point-value reconstruction if the scheme is TVD. When the scheme is not TVD, point-value reconstructions have some advantage in performance.

From the basis of the work given here, the design of high-resolution upwind shock-capturing methods can be advanced in a more unified manner. This should yield benefits for all of the methods falling into this general category.

Preface

The path which led me to this point is worth exploring before going further. My interest in numerical methods for fluid flow began as a prerequisite for the effective modeling of heat pipes. I began by studying the work of S. V. Patankar [1] as suggested by Dr. D. V. Rao. Over time, I became somewhat displeased with the nature of the methods, their results, and limitations. By this time, I had become interested in the numerical methods for fluid flow as things unto themselves. This time coincided with my beginning employment at Los Alamos National Laboratory. Shortly before arriving to work at the lab, I began to be interested in the work of Dennis Liles [2] for modeling two-phase flows. This work is based in large part on the earlier work of Harlow and Amsden on the ICE methods [3].

While investigating these methods, I came across a book by Oran and Boris [4]. The viewpoint expressed there was different than anything I had looked into before and I found the methodology intriguing to say the least. Initially, I was very impressed by the flux-corrected transport methods described by Oran and Boris when compared to the classical methods I was used to. When I tried to use these methods on a more complex, system of equations problem, I saw a number of problems with the solutions. These observation formed the genesis of the research that followed.

Soon, I began to read and attempt to understand total variation diminishing schemes and later high-order Godunov methods. Both of these method types were similar to the flux-corrected transport, but their performance on systems of equations is significantly better. They seemed to have a much more appealing mathematical basis. It was seeking the answer to the questions: how can flux-corrected transport be improved? and how are flux-corrected transport, total variation diminishing and high-order Godunov methods related? that produced Chapters 5, 6 and 7.

Further work presented here primarily centered about answering several questions about the use of high-order Godunov methods. The ties made in Chapter 6 makes this applicable to the other method categories mentioned here. Chapters 7 and 8 expand the line of thought taken with the flux-corrected transport methods and look at the problem of designing limiters for second-order high resolution schemes. Limiters are at the core of the construction of this type of numerical method and understanding them is essential. The last two chapters of the dissertation clean up loose ends. Chapter 9 addresses some questions in reconstruction methods for high-order Godunov methods.

This dissertation can be viewed as a skewed reflection of my own evolution in the understanding of these methods. I started by looking at FCT methods and ended up relying on HOG methods for algorithm design. The reason for this is that the Godunov-type methods are more physically and mathematically (philosophically) appealing to me. This is a matter of personal taste, but I do believe that they represent an effective basis for future development along a number of fronts.

The work found in [5] has been accepted for publication in *Communications in*

Applied Numerical Methods. This work forms part of Chapter 8.

Finally, the bulk of the work presented in this dissertation has been submitted in the form of papers to several professional journals. References to these can be found in the bibliography [6, 7, 8, 9, 10, 11].

Notation

The notation used in this work requires a short explanation.

References are denoted by square brackets. Therefore the third reference would be seen as [3]. The references are listed in order of their use. When more than one reference is given, the first reference is the recommended one.

Equations are denoted by regular parenthesis. The fourth equation in the sixth chapter is referenced by (6.4).

Theorems and similar structures will be referenced if their proofs exist in the literature. Those proven by myself will not contain a reference with their labels.

Table of Contents

List of Figures

List of Tables

Chapter 1.
Overview

> Although a meal can be enjoyed without understanding the process of digestion, numerical methods should be both understood and enjoyed. This requirement is not merely the whim of a tidy mind, for a method once understood can often be improved with little effort. *J. J. Monaghan [12]*

The topic of this dissertation is the design of high-resolution upwind shock capturing methods. By high-resolution I mean that the method is capable of resolving various fine detail features of the solution field without resorting to an excessively fine grid. Upwind makes reference to the method's use of the mathematical/physical structure of the solution field, and the governing equations in constructing the numerical method. Finally, the adjective shock-capturing clarifies the type of method developed. Some methods track discontinuities or shocks in the solution field and essentially use these tracked features as internal boundaries. Shock-capturing methods do not do this, and "capture" discontinuities without modification of the method used throughout the solution field.

The next three chapters give a brief introduction to these topics. The first of these three chapters gives background and motivational information regarding the study of this topic. Classical shock-capturing methods are the topic of the second of these chapters. These classical methods provide the foundation for the work that follows. The third and final introductory chapter gives an introduction to modern high-resolution shock-capturing methods, and the categories they fall into.

Following this introduction, I introduce the topic of method design. This begins with the method known as flux-corrected transport (FCT). The FCT method is known to have certain pathological problems, and this chapter addresses this matter in a systematic fashion. Through this analysis it becomes clear that the FCT is more intimately related to other modern methods, most notably symmetric total variation diminishing (TVD) methods. This relation is expanded upon and exploited in improving the FCT method's performance. In the chapter that follows, the combined FCT/Symmetric TVD methods are related more closely to high-order Godunov (HOG) methods. The HOG methods are a philosophically satisfying means of defining high- resolution upwind shock-capturing methods because the process is divided into two parts: reconstruction (interpolation) and evolution (upwinding). This decoupling of the method development allows one to concentrate on one or the other feature. From this, unity of the methods is demonstrated, and new, improved methods can be derived.

The two chapters following this unification of the methods discuss the construction

of limiters. Limiters are the means through which modern methods are differentiated from classical methods. Their construction is the most important portion of method design, and has a profound impact on a method's performance. Past studies of limiters have been narrowly focused, and these chapters are aimed at broadening this view. Finally, a chapter on some basics of the reconstruction step are discussed with a critical view taken of current practices.

In the appendices a number of more practical aspects of extending these methods to systems of equations are discussed.

Chapter 2.
Introduction

Of a good beginning cometh a good end. *John Heywood*

2.1 Background and Motivation

Recently, several articles have appeared highlighting the importance of numerical approximation of conservation laws from both a theoretical and practical standpoint [13, 11]. High quality numerical approximations to conservation laws are necessary in a number of endeavors as noted at the end of this chapter. Numerical work is also becoming increasingly important for theoretical studies. In a very real sense, numerical experimentation is becoming a third major thrust of science along side experimental and theoretical work.

This chapter gives an introduction to the subject of numerical approximations to hyperbolic conservation laws (HCLs). It covers the basis and motivation for the study of the subject and provide a brief introduction to some of the important theoretical concepts in Section 2.2. Also, the basic philosophy used in developing numerical algorithms to solve these sorts of equations is presented in Section 2.3. A number of applications of the accurate solution to HCLs is presented in Section 2.4. This serve to underline the importance of this subject to a wide range of scientific pursuits.

The primary motivation for pursuing any subject is to seek understanding. In a number of diverse fields, a similar process is responsible for a rich variety of physical (or mathematical) behavior. The role of transport of some quantity (like mass, energy, particles, sound, wave packets etc.) can be thought of to me at the heart of most physical processes (the last section of this chapter contains a longer more detailed list). These physical systems can all be characterized at a simple level by the same model equation,

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0 . \tag{2.1}$$

where u is the transported quantity and $f(u)$ is the flux function for this quantity. In mathematical terms, this is a hyperbolic equation if $\partial f/\partial u$ is a real number. This equation describes the transport and conservation of u in the $x-t$ plane. In general, this equation can represent a system of equations as well. In that case, u and $f(u)$ are vectors. Section 2.2.1 covers this subject in more detail. Thus, (2.1) represents the basic for of a HCL.

The solution of the above equation exists in closed form for only a few simple, idealized cases thus some approximations must be made to solve it in the general case.

If the approximations are sufficiently detailed and accurate, the solutions found can exhibit the wide range of nonlinear behavior and rich phenomena found in nature. As is discussed later, good approximations can also lead to the discovery and/or clarification of physical phenomena [15, 16].

A number of detailed references on these subjects exist in the literature. On the basics of HCLs some of the recommended references are Lax [17, 18], Smoller [19], Landau and Lifshitz [20], Mihalas and Mihalas [21], Duderstadt and Martin [22], Chorin and Marsden [23], Anderson [24] and Courant and Friedrichs [25]. These references present the material in a readable informative manner, although they vary in emphasis and difficulty. All of these references are biased in the direction of fluid flow (except Duderstadt and Martin), but considering that that is the most common application, this is understandable.

From the presentations found in both Mihalas and Mihalas and Duderstadt and Martin it can be seen that fluid equations can be viewed as continuum extensions of the Boltzmann transport equation (via a Chapman-Enskog expansion or similar procedure). The Boltzmann transport equation has a form that is very similar to (2.1) [26, 27]

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f = S_{coll} \,, \tag{2.2}$$

where f is a time dependent distribution function, $f(\mathbf{r}, \mathbf{u}, t)$, in position and velocity space. S_{coll} is a scattering kernel that I ignore. In fact, with S_{coll} set to zero, the equation is the multidimensional equivalent to (2.1) with constant velocity by setting $\rho = f$. Additionally, the diffusive terms in the full set of equations (Navier-Stokes) can be viewed similarly. This "transport" viewpoint has been an active area of research in hyperbolic heat conduction [28, 29]. Similar lines of thought can be found in radiation transport in the passage from a transport to diffusive approximation to the Boltzmann transport equation [22].

Remark 1 ***The solution collisionless to the Boltzmann equation is explored in some depth by Harten, Lax and van Leer [30] with relation to the general solution of HCLs. This has specific application to a method known as flux splitting which is covered in some detail in Appendix B.***

The numerical solution of equations of this sort (for continuum approximations) can be found in a number of sources as well. The most basic and perhaps elegant source is Richtmyer and Morton's book [31] which contains much of the basic theory to support classical methods of solution. The history of computational fluid dynamics (CFD) is presented in Roache [32], Potter [33] as well as Anderson, Tannehill and Pletcher [34]. Roache contains a complete account of the early development of CFD and a large number of references. More recent developments are covered in several texts: Oran and Boris [4], Hirsch [35, 36] and Fletcher [37, 38]. The text by Oran and Boris is especially recommended as an introduction to the entire subject of numerical

solution of complex physical problems as well as HCLs. A book by Sod [39] contains a good deal of mathematical theory. Recently, LeVeque has released some lecture notes in the form of a monograph [40]. This work is highly recommended as an introduction to conservation laws from both a mathematical and numerical perspective. In addition to these books, a number of survey papers have appeared in recent years; these include [41, 42, 43, 44, 45, 46, 47, 48]. An interesting survey of methods has been done in relation to nonlinear acoustics of rocket engines [49] as an extension of the review by Baum and Levine [50]. This survey underlines the point that fewer and fewer approximations are necessary in the analysis of physical systems because of the power of modern hardware and algorithms.

Figure 2.1 shows the rough family tree for the development of upwind (explained later) approximations to (2.1). Beginning with the work of Richardson [51] on the solution for stress in dams and moving on to the paper on partial differential equation by Courant, Friedrichs and Lewy [52] this subject had its genesis Von Neumann and Richtmyer [53] introduced artificial viscosity which was followed shortly by two methods that did not introduce numerical dissipation artificially [54, 55], but did through the nature of the finite difference equations. The beginnings of more powerful methods for solving HCLs can be found in several papers by Godunov [56, 57] and Lax and Wendroff's famous paper [58]. These papers lead to several seminal works by Boris and Book [59] and van Leer [60] who were the first to recognize the importance of nonlinearity in difference schemes. These two papers were at the root of a large set of work in the last twelve years highlighted by the work of Harten [61], Zalesak [62], Roe [63] and a group of researchers at UCLA [64, 65, 66] where the earlier work was clarified and extended. It is the construction of these approximations that is the subject this research topic.

2.2 A Mathematical Introduction

Consider the same equation as above

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0\,, \tag{2.3a}$$

which is a first-order hyperbolic transport equation for u and as before f is the flux of u. Equation (2.3a) can be written as

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0\,, \tag{2.3b}$$

where the flux Jacobian is defined by

$$a = \frac{\partial f}{\partial u}\,.$$

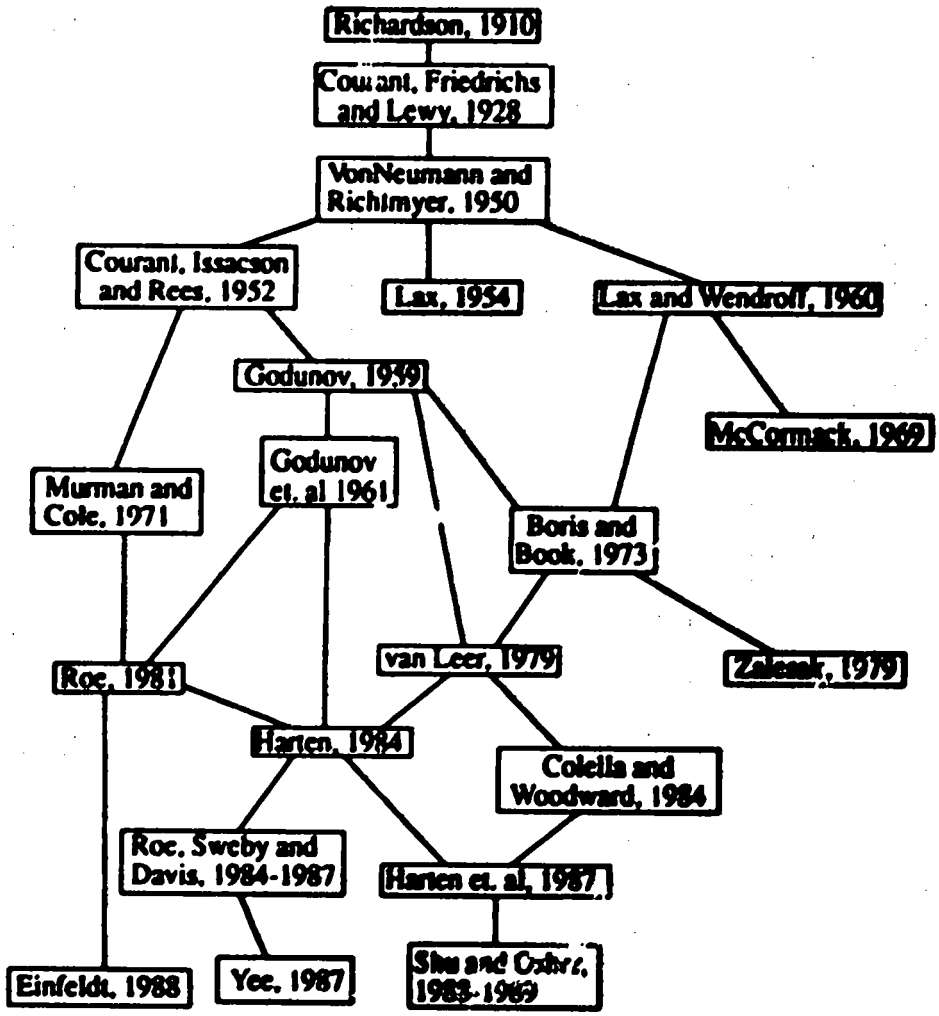


Figure 2.1: This shows a rough genealogy for computational fluid dynamics using upwind discretization methods.

If the characteristic speed, a, is constant for all x, then an exact solution exists for (2.3b). This solution is

$$u(x,t) = u_o(x - at) \ , \tag{2.4}$$

where $u_o(x) = u(x,0)$ is the initial condition. This defines the scalar wave (Kriess) equation. For a more general prescription of f a closed form solution does not exist.

2.2.1 Systems of Hyperbolic Conservation Laws

A system of m conservation hyperbolic laws can be similarly defined; however, the behavior which it describes is considerably more complex. Consider

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = 0 \ , \tag{2.5a}$$

which is a set of hyperbolic conservation laws where $\mathbf{U}$ is a column vector $(u^1, u^2, \ldots, u^m)^T$ of conserved quantities and $\mathbf{F}$ is a column vector $(f^1, f^2, \ldots, f^m)^T$ of fluxes of $\mathbf{U}$. Equation (2.5a) can be written as

$$\frac{\partial \mathbf{U}}{\partial t} + A\frac{\partial \mathbf{U}}{\partial x} = 0 \ , \tag{2.5b}$$

where

$$A = \frac{\partial \mathbf{F}}{\partial \mathbf{U}} = \begin{bmatrix} \partial f^1/\partial u^1 & \ldots & \partial f^1/\partial u^m \\ \vdots & \ddots & \vdots \\ \partial f^m/\partial u^1 & \ldots & \partial f^m/\partial u^m \end{bmatrix} .$$

The matrix A is the flux Jacobian for the system defined by (2.5b).

In general, equations of the type considered above can develop discontinuous solutions even when the initial data is smooth. Because of this, the solutions are not unique. To rectify this, the admissible solutions must satisfy an entropy condition (for details on this see [17, 18, 19, 40] see [67] for a simple introduction). It is the formation of discontinuities in the solution that causes the difficulties for finite-difference solutions of (2.3b). At these discontinuities, the function ceases to be smooth, and the usual assumptions made in constructing finite- difference approximations collapse. As a result, more physical information needs to be incorporated into the solution procedure.

The system of equations is classified as hyperbolic if all the eigenvalues of A are real [30]. These eigenvalues λ_k can be arranged in the order of increasing magnitude, thus

$$\lambda_1 < \lambda_2 \ldots < \lambda_k < \ldots \lambda_{m-1} < \lambda_m \ .$$

Lax [18] has defined entropy conditions for hyperbolic equations and systems. Given

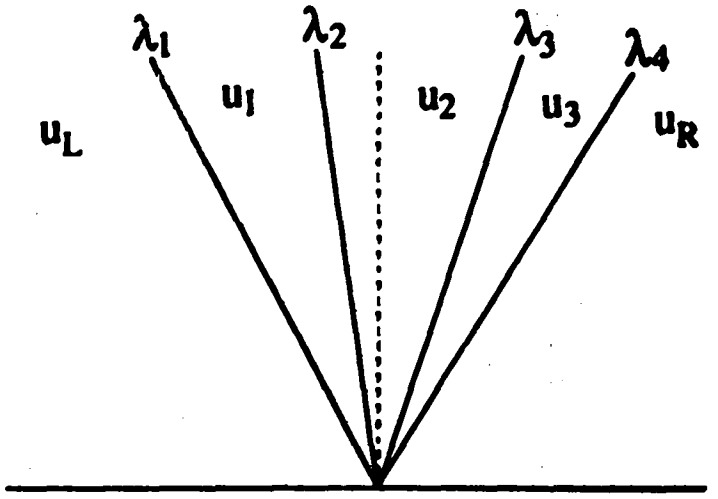


Figure 2.2: The left and right states have m waves associated with them (4) in this case and $m-1$ constant states between them for $t > 0$.

two states u_R and u_L at $t = 0$ (in one spatial dimension), which exist to the right and left of a discontinuity respectively, the admissible speed of the discontinuity must satisfy this inequality

$$\lambda(u_L) > s > \lambda(u_R) \ , \tag{2.6a}$$

where s is the speed of the discontinuity. For systems this condition is

$$\lambda_k(u_L) > s > \lambda_k(u_R) \ , \tag{2.6b}$$

with

$$\lambda_{k-1}(u_L) < s < \lambda_{k+1}(u_R) \ . \tag{2.6c}$$

These conditions form an entropy condition for systems. Stated in other terms, this means that the entropy must either remain constant or increase in a system. An increase in entropy occurs across discontinuities. These conditions must be met for a solution to the system to be physical in nature. Menikoff and Plohr [68] explore more general cases. In some cases especially near phase transitions, the isentropes of the system fail to be convex thus causing physical solutions to violate Lax's conditions.

Lax also states that for a system of m equations, $m-1$ constant states exist between the left and right states at $t > 0$ (see Fig. 2.2). These states can be separated by rarefaction or shock waves or contact discontinuities. A rarefaction is a smooth expansive transition, while a shock is a sharp sudden change where the flow is discontinuous. A contact discontinuity is like a shock, but some quantities may be continuous across it.

An additional manner of characterizing systems (or equations) of HCLs is to analyze the structure of the eigenvalues. Lax [69, 17, 18] defines an eigenvalue as

being linearly degenerate if

$$\frac{\partial \lambda_k}{\partial \mathbf{U}} \cdot \mathbf{r}_k = 0 \,, \tag{2.7a}$$

and as genuinely nonlinear if

$$\frac{\partial \lambda_k}{\partial \mathbf{U}} \cdot \mathbf{r}_k \neq 0 \,, \tag{2.7b}$$

where $\mathbf{r}_k$ is the k^{th} right eigenvector. An example of a linearly degenerate eigenvalue is the characteristic speed in the scalar wave equation ($\lambda_1 = a$, $\partial a/\partial u = 0$, and $\mathbf{r}_1 = 1$). A genuinely nonlinear eigenvalue can be found in Burger's equation ($\lambda_1 = u$, $\partial u/\partial u = 1$, and $\mathbf{r}_1 = 1$). These equations can thus serve as models for the behavior of these types of waves in more complex equation(s). In the Euler equations (discussed in detail in Appendix B) the eigenvalues associated with sound waves are genuinely nonlinear while the eigenvalue(s) associated with fluid motion is linearly degenerate. A shock is associated with genuinely nonlinear eigenvalues while a contact discontinuity is associated with linearly degenerate eigenvalues. A shock in this sort of system is referred to as a k-shock and a rarefaction as a k-rarefaction. For contact discontinuities, the above relations must be modified to read

$$\lambda(u_L) = s = \lambda(u_R) \,, \tag{2.8}$$

thus the flow speed remains constant across the contact discontinuity.

Remark 2 *Systems of conservation laws which are not strictly hyperbolic [70] has been the subject of intense research lately. This is a topic of theoretical and practical interest which has direct application to three-phase flow in porous media which form a two equation system of conservation laws in one dimension. The numerical solution of such systems is following suit and benefiting greatly from the recent increase in theoretical understanding. Another related area that could benefit from some theoretical/numerical work is two-phase flow [71, 72]. The application of numerical methods to two-phase flow has a number of striking similarities to multiphase flow in porous media.*

These equations admit discontinuous solutions thus requiring that the solution converge in a weak rather than a strong sense. By a weak solution I mean that solutions satisfy (2.1) in the sense of distributions [30, 19], i.e.,

$$\int_0^\infty \int_{-\infty}^\infty \left[\frac{\partial \phi}{\partial t} u + \frac{\partial \phi}{\partial x} u \right] dx\, dt + \int_{-\infty}^\infty \phi(x,0)\, u_0(x)\, dx = 0 \tag{2.9}$$

for all C^∞ test functions $\phi(x,t)$ with compact support.

The above statement can be reformulated to give a form useful for the construction

of difference schemes. Integrating (2.1) over the rectangle $(x_0, x_1) \times (t_0, t_1)$ gives

$$\int_{x_0}^{x_1} u(x,t_1)\,dx - \int_{x_0}^{x_1} u(x,t_0)\,dx + \int_{t_0}^{t_1} f(u(x_1,t))\,dt - \int_{t_0}^{t_1} f(u(x_0,t))\,dt = 0\,. \quad (2.10)$$

Thus where the solution is smooth, (2.1) holds, but across curves of discontinuity, the Rankine-Hugoniot condition holds as

$$s(u_R - u_L) = f(u_R) - f(u_L)\,, \quad (2.11)$$

where s is the speed of the discontinuity and u_R and u_L are the states to the left and right of the discontinuity, respectively. For numerical work the above statement is quite profound. The solutions are conserved cell-averages rather than point-values and the fluxes are time averages of the flux at the cell boundaries. These definitions are convenient for use with finite volume discretizations.

It is well known that the weak solutions to (2.1) are not unique. To find the correct solutions, an additional condition must be met. This type of condition is known as an entropy condition after the physical quantity of the same name [17, 18]. In [73], it was shown that entropy satisfying solutions of (2.1) are limiting solutions to a parabolic equation

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = \epsilon \frac{\partial^2 u}{\partial x^2}\,, \quad (2.12)$$

with $\epsilon > 0$ and the limit being taken as $\epsilon \downarrow 0$. This connection is explored at some length in Chapter 8.

2.2.2 The Rankine-Hugoniot Conditions

The Rankine-Hugoniot conditions are especially important to the theory of conservation laws when solutions are discontinuous. Several elegant proofs are available in the literature. One is found in [18]. Referring to Fig. 2.3 and defining

$$U(t) = \int_a^b u(x,t)\,dx = \int_a^y u(x,t)\,dx + \int_y^b u(x,t)\,dx\,, \quad (2.13a)$$

differentiating with respect to time, and using the governing equation (2.1) one gets

$$\frac{dU}{dt} = \int_a^y \frac{\partial u}{\partial t}dx + u_L s + \int_y^b \frac{\partial u}{\partial t}dx + u_R s\,, \quad (2.13b)$$

where u_L and u_R are the states to the left and right of the curve of discontinuity $x = y(t)$ and $s = dy/dt$. Using $\partial u/\partial t = -\partial f/\partial x$ and carrying out the integration one gets

$$\frac{dU}{dt} = f_a - f_L + u_L s - f_b + f_R - u_R s\,. \quad (2.13c)$$

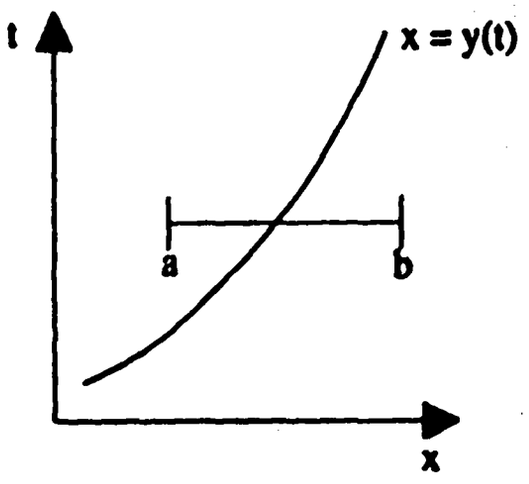


Figure 2.3: A pictorial representation of the domain used in the proof of the Rankine-Hugoniot condition (adapted from [18].)

with the conservation law stating that

$$\frac{dU}{dt} = f_a - f_b \,, \tag{2.13d}$$

then

$$s[u] = [f] \,, \tag{2.13e}$$

where $[u] = u_R - u_L$ and $[f] = f_R - f_L$. In [23] another proof is given

2.3 General Numerical Philosophy

This section covers the basic philosophy used in the numerical approximation of HCLs. The methods discussed here can all be classified as finite difference or finite volume type methods [74]. Because of Lax and Wendroff's theorem [58] concerning the nature of solutions to HCLs, the equations are always differenced in conservation form.

Theorem 1 (Lax and Wendroff [58]) *If a difference equation is in conservation form and is consistent with the original conservation law as well as stable, it converges to a weak solution of that conservation law.*

With this form, the solutions converge to solutions which satisfy the Rankine-Hugoniot conditions. Conservation form implies that quantities are conserved numerically, as they are physically, thus when a domain is subdivided into a set of subdomains (control volumes), the amount of material exiting one subdomain exactly enters the subdomain adjacent through a common interface.

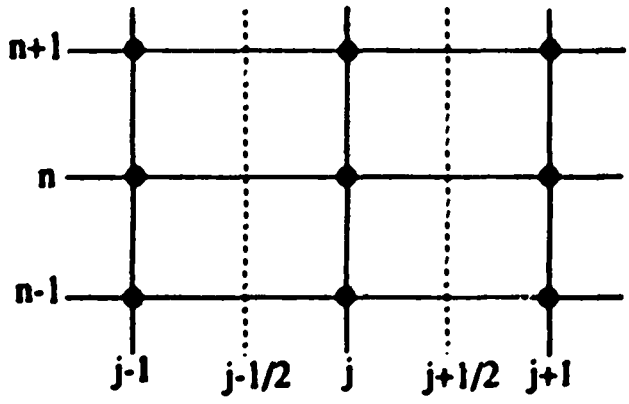


Figure 2.4: The spacetime grid is shown with the grid interfaces denoted by the dotted lines and the computational nodes by the dark circles.

These schemes can be expressed in the following form,

$$u_j^{n+1} = u_j^n - \sigma\left(\dot{f}_{j+\frac{1}{2}} - \dot{f}_{j-\frac{1}{2}}\right) = 0\,, \tag{2.14a}$$

in one dimension where $\sigma = \Delta t/\Delta x$, or more generally

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{V_j}\sum_k^N A_k \dot{f}_k = 0\,, \tag{2.14b}$$

for a homogeneous governing equation. Here j refers to the index of a control volume, n to the time level (see Fig. 2.4), and $\dot{f}$ is the numerical flux. In the general case V_j is the cell volume, A_k is the area of a face of that volume with a total of N faces (sides) to a volume. The above equations can also be written in a semi-discrete form

$$\frac{\partial u}{\partial t} = -\frac{1}{V_j}\sum_k^N A_k \dot{f}_k = 0 \tag{2.14c}$$

The determination of the numerical fluxes, $\dot{f}$ is at the heart of the subject. To insure that the solutions are consistent then

$$\dot{f}(u, u, \ldots, u) = f(u)\,. \tag{2.15}$$

Given this condition with the stability of the overall solution procedure implies the convergence of the scheme by the Lax equivalence theorem [31, 67].

Theorem 2 (Lax equivalence theorem) *Given a well-posed initial value problem and a corresponding numerical approximation that is consistent, then stability is a necessary and sufficient condition for (equivalent to) convergence.*

Unfortunately, this theorem can only be applied to the *linear* types of schemes like those described in Chapter 3. Nevertheless, this theorem is important and can be used to analyze linear methods that are the building blocks of more advanced methods.

As a measure of the accuracy of the solution, I use the Taylor series to act as a measure. This means that if a method is stated to be r^{th} order accurate, the leading term in the truncation error is $\mathcal{O}(\Delta x^{r+1})$. Later, the problems associated with this are discussed. In general, the general Taylor series driven difference approximations are used in favor of a polynomial approximation driven approximations (although Taylor series are often used to measure the polynomial's accuracy). This is motivated by the course of recent developments in numerical algorithms for solving HCLs. One caveat with the use of Taylor series based measures of accuracy is that discontinuities make the concept of accuracy somewhat meaningless at those points.

The accuracy of solutions can also be measured in terms of norms. The three most commonly used norms are the L_1, L_2 and L_∞ (also known as the maximum) norms. These are defined by

$$L_1 = \sum_j^N \frac{|e_j|}{N} , \tag{2.16a}$$

$$L_2 = \left(\sum_j^N \frac{e_j^2}{N} \right)^{\frac{1}{2}} , \tag{2.16b}$$

$$L_\infty = \sup (|e_j|) , \tag{2.16c}$$

where

$$e_j = U_j^{exact} - U_j^{approx.} ,$$

given an exact solution. Although this gives a quantitative measure of algorithm performance, the qualitative measure of performance is also generously used. These two means of measure should provide a complementary means of determining solution qualities.

2.4 Applicability to Other Disciplines

The successful solution HCLs is vital to a large number of endeavors. This general problem is present in any system where fluid flow is present (with the exception of Stokes flow or subsonic potential flow, but these represent simplifications of the actual physical system). Thus the range of applicability is quite large. The methods discussed in the next chapter have been found to be useful in the solution of aerodynamic flows [43, 45, 36, 75] where they are currently widely used. These methods (the modern advection solution algorithms) are also finding use in turbulence modeling. The process of large-eddy simulation [76] involves the solution of fluid equations with only the large (kinetic energy carrying) structures being resolved. Recently, it has

been proposed that modern advection algorithms (see Chapter 4) could serve as a turbulence model [77, 78, 79].

Methods of a modern type are also finding use in the solution of incompressible flows (the flows above are primarily compressible). The solution of this type of problem is largely dominated by first-order schemes [1], but recently second and third order methods have become more widely used [80]. Leonard [81, 82, 83, 84] has developed a scheme based on his QUICK[1] scheme, which has a great deal in common with some other modern algorithms. This method or one like it has the promise of greatly improving codes currently used to compute a variety of industrial flows. Other workers have also applied other modern methods to more classical incompressible flow solvers [85, 86, 87].

The solution of these equations is also very useful in astrophysical fluid dynamics [88, 89, 78, 90, 91, 92]. The physical systems in astrophysics place severe demands on numerical methods [27], and the methods must be carefully designed to compute solutions with needed accuracy. Other flows of a geophysical nature are amenable to modern approaches to solving advection [46, 93, 94].

The solution of wave equations is important in applications which use a fully Lagrangian formulation [95]. In these methods, the grid flows with the fluid thus leaving only sound waves explicitly in the equation set. The solution of this sort of system is amenable to similar methodology as other wave equations. The Lagrangian formulation often rids the problem of the linearly degenerate eigenvalue(s) (they go to zero), but still leaves genuinely nonlinear eigenvalues in the set. Thus the primary approximation problem still exists.

As mentioned earlier, the hyperbolic heat conduction problem is open to numerical solution by methods applicable to HCLs. The quality of the solution is significantly enhanced through the use of modern algorithms [96]. Also mentioned earlier was the work of Brio and Wu [15], which solved the MHD equations. Using modern algorithms new phenomena were discovered, which may have been validated by observations [16]. Also along these lines is the solution of problems in electromagnetism by methods developed for compressible aerodynamics [97, 98] with promising results.

Several uses in nuclear engineering applications requiring thermal hydraulic analysis can be found in [99, 100, 101]. These methods are also showing a great deal of use in the modeling of solid dynamics under severe physical conditions [102] where the solid behaves in a fluid-like manner. Additional applications can be found in reservoir modeling [103, 104, 105] with implications to petroleum recovery.

In the next chapter I explore some of the classical numerical methods for solving conservation laws and the problems associated with them.

[1]The QUICK scheme uses a third-order (spatially) upwind algorithm based on a finite difference stencil containing the one downwind point and two upwind points. It can also be derived by means of quadratic polynomials

Chapter 3.
Classical Methods for Conservation Laws

> The present contains nothing more than the past and what is found in the effect was already in the cause. *Henri Bergson*

3.1 Introduction

In this chapter, several of the most important classical methods for solving HCLs is covered. These methods although outdated by modern standards still comprise the backbone of most modern methods, and contain some of the essential concepts for the successful design of numerical schemes. This chapter discusses the basic construction of these methods, their stability and other pertinent properties.

Errors in the numerical solution of hyperbolic problems are generally classified as being of either a damping or a dispersive variety. As is seen below, a useful numerical scheme must contain some minimal amount of dissipation to remain stable and produce physical solutions. This dissipation damps out error which would otherwise grow in an unbounded fashion, but it also destroys many features of the flow field. Lack of sufficient damping results in dispersive errors that can cause unphysical maxima and minima to be created in the solution by the numerical scheme.

Phase errors result in information being transported at a numerical velocity below or above the true velocity of this information. These errors are depicted in Fig. 3.1. Typically, VonNeumann stability analysis [31, 4, 35, 37] is used to analyze these errors. The process consists of replacing the dependent variables by Fourier series, $e^{ijm\theta}$, defining the new time value of the variable to be equal to Fourier series at the old time multiplied by a function A or the amplification factor, in general

$$u_j^{n+1} = g\left(u_k^n, u_k^{n+1}\right) \Rightarrow Ae^{ijm\theta} = g\left(e^{ikm\theta}, Ae^{ikm\theta}\right) . \tag{3.1}$$

Generally, the expression of A is a combination of real and imaginary trigonometric terms and is transformed to extract useful information. This is accomplished by separating the functional form of A into two pieces: an amplification factor and a phase angle,

$$A(k\theta) = |G|e^{i\phi} , \tag{3.2}$$

where G is the magnitude of A and $\phi = \tan^{-1} \mathrm{Im}(A)/\mathrm{Re}(A)$ is the phase angle. For stability, G must be less than or equal to one for all $k\theta$, but this implies damping. Small values of G imply excessive damping. For the scalar wave equation, the exact

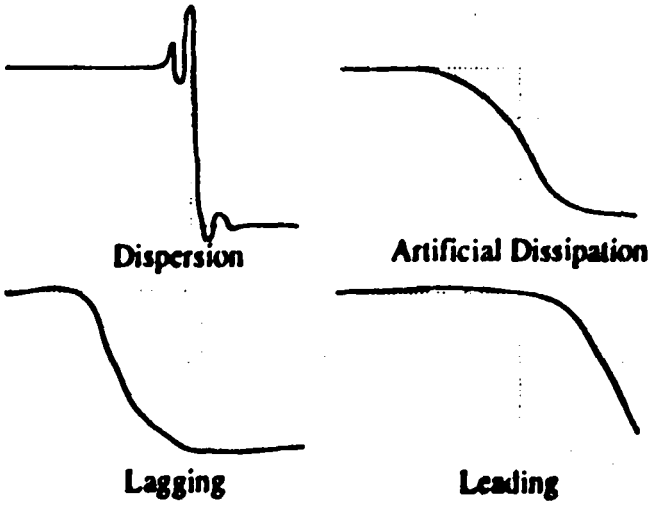


Figure 3.1: Here the three main types of errors in the solution hyperbolic initial value problems are shown: artificial dissipation, dispersion, leading and lagging phase errors. (The exact solution is in the lighter pen and the representation of the numerical solution is in the darker pen.)

phase speed is known[1] so that the ratio of this to the numerical phase speed can be taken. If this quantity is less that one the error is lagging, if it is greater than one the error is leading (see Fig. 3.1). These errors have a spectrum of values which can have a large range of values.

All the methods discussed in this chapter are explicit in nature and are thus limited by a stability condition (some multiple of the Courant-Friedrichs-Lewy (CFL) [52] number). This number, $\nu = |a|\,\Delta t/\Delta x$, is a dimensionless value which describes the proportion of the domain of dependence covered during a time step (see Fig. 3.2). These methods are: the central difference method with or without artificial diffusion, upwind differencing, the Lax-Friedrichs method, the Lax-Wendroff method, and the Beam-Warming scheme or second-order upwind differencing.

3.2 Central Differencing and Artificial Diffusion

The simplest type of numerical scheme seems to be a very natural manner to deal with the hyperbolic equation. This method deals with approximating the first derivative of the flux function with a centered spatial difference which has second-order accuracy and marching explicitly in time and is known as the forward time-centered space (FTCS) scheme. This method can be written

$$u_j^{n+1} = u_j^n - \frac{\sigma}{2}\left(f_{j+1}^n - f_{j-1}^n\right) , \tag{3.3}$$

[1] The exact wavespeed is $\nu\phi$ where ν is the CFL number.

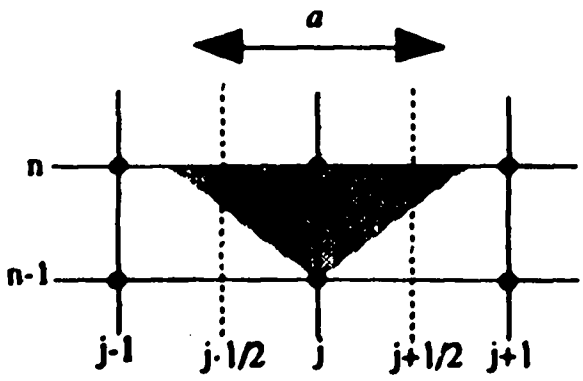


Figure 3.2: An interpretation of the CFL limit sketched in the $x - t$ plane for point j. For an explicit calculation, information should not be transported more than one mesh interval from its origin or in other words the adjacent grid points must lie on or outside the domain of dependence ($\Delta x \geq a\Delta t$). If waves from two different grid points are not allowed to interact, the restriction becomes twice as severe.

where $\sigma = \Delta t/dx$ for uniform grid spacing. This is equivalent to saying that the cell edged fluxes are the arithmetic mean of the neighboring grid points or taking the fluxes to be a linear interpolation of the initial data. Thus the numerical flux functions are

$$f^n_{j+\frac{1}{2}} = \frac{1}{2}\left(f^n_j + f^n_{j+1}\right) . \tag{3.4}$$

Unfortunately, this method can be shown to be unconditionally unstable, with errors growing in an unbounded manner. This behavior can be seen in Fig. 3.3 plotted after 20 time steps showing the impending disaster.

Through the addition of artificial dissipation [53, 31] this solution method can be resurrected to some degree. This requires the addition of a term on the right hand side of the equation which acts in the same fashion as physical dissipation. The coefficient is somewhat arbitrary, but too little dissipation results in a more stable, but low quality solution. Too much diffusion[2] can either result in destroying some or all of the features of the solution or causing a new instability because of the stability restriction implied by the explicit diffusion equation. Results using the FTCS scheme with artificial dissipation are shown in Fig. 3.4. The dissipation has largely cured the instability, but now the solution exhibits a large leading phase error. Smarter forms of artificial viscosity are used (see Jameson [106] for example) with acceptable performance, but the methods are always somewhat *ad hoc* in nature [107].

[2]The terms diffusion and dissipation are used interchangeably in the text. They should be treated as synonyms

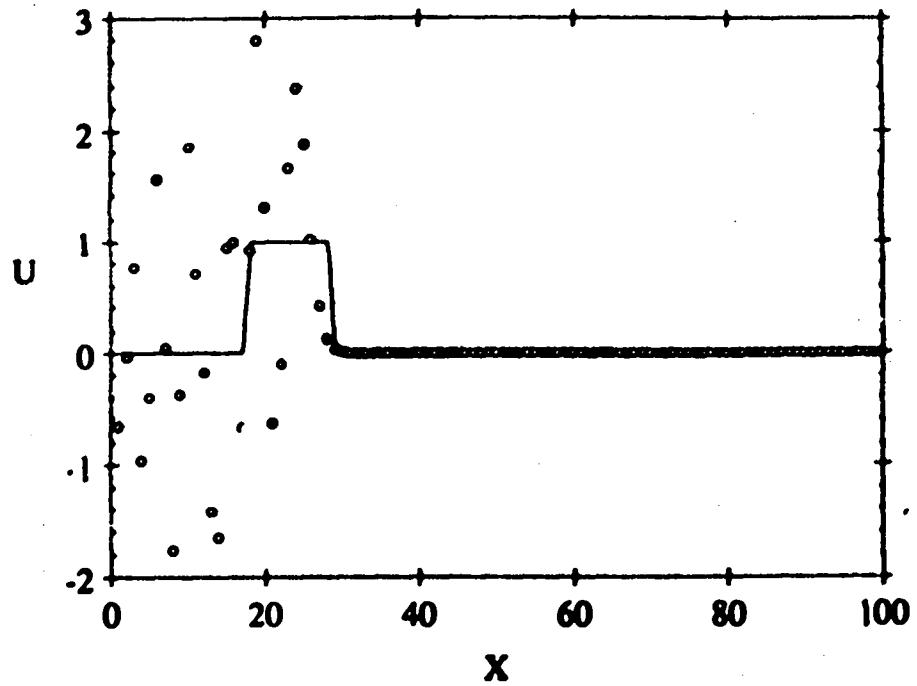


Figure 3.3: The results found using the FTCS scheme show the growth of instabilities and their unbounded growth. (The exact solution is in the solid pen and the numerical solution is denoted by the circles.)

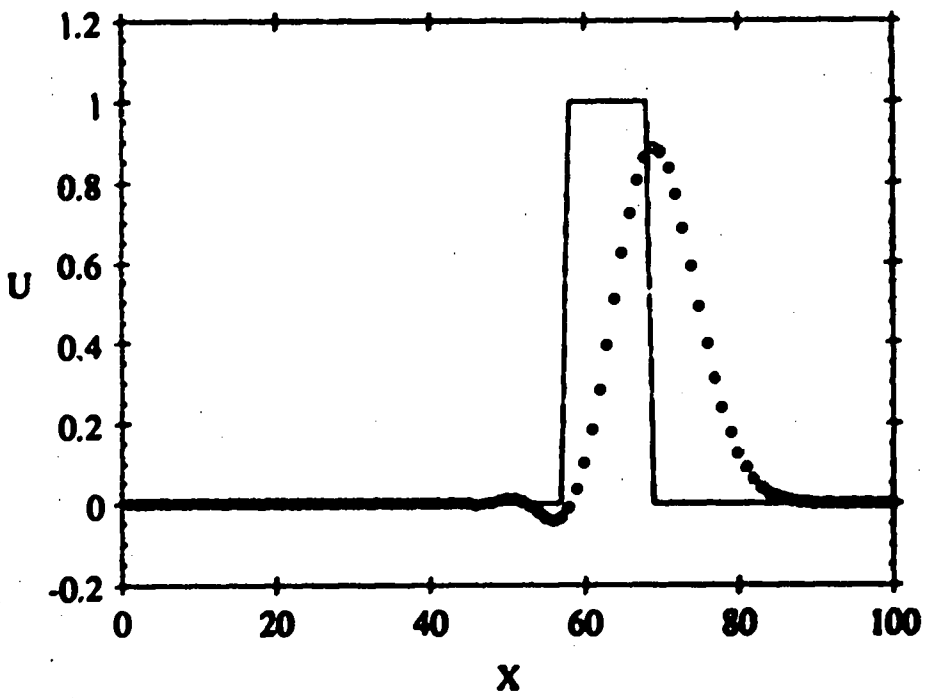


Figure 3.4: The results found using the FTCS scheme with an artificial dissipation coefficient of 0.1 ($a = 1$ and $\nu = 0.5$).

3.3 Upwind Differencing Type Methods

The behavior discussed in the last section is clearly unacceptable although useful computations can be performed using artificial diffusion because it does converge to the correct solution [108]. In [54] a new more physically based approximation is described. This method forms the basis for a large class of modern numerical methods in Chapter 4 (see Fig. 2.1).

This method is first-order accurate in both time and space, and takes the direction of the wave propagation in the problem into account when computing the cell-edge fluxes. There are several ways to derive this approximation, which all have relative advantages. Typically, this scheme can be derived with a first-order Taylor series approximation which is biased by the direction of the flow locally. This results in a difference scheme for (2.1) like

$$u_j^{n+1} = u_j^n - \sigma a \left(u_j^n - u_{j-1}^n \right) , \tag{3.5}$$

where $a > 0$, this can also be written in conservation form by stating

$$f_{j+\frac{1}{2}} = a u_j^n .$$

Another way to write the cell-edge fluxes is [109]

$$f_{j+\frac{1}{2}} = \frac{1}{2} \left[f_{j+1}^n + f_j^n - |a| \left(u_{j+1}^n - u_j^n \right) \right] , \tag{3.6}$$

where $f_j^n = a u_j^n$. This form is advantageous because it shows the magnitude of the diffusion associated with the spatial differencing. For the upwind differencing, the numerical diffusion coefficient is

$$d^{upwind} = |a| \frac{\Delta x}{2} . \tag{3.7a}$$

The effective induced viscosity is

$$d^{upwind} = \frac{|a| \Delta x}{2} (1 - \nu) , \tag{3.7b}$$

which reflects the fact that the upwind differencing recovers the exact solution to the scalar wave equation if $\nu = 1$ [30]. This term can be determined from the comparison of upwind differencing with the FCTS scheme assuming the Lax-Wendroff scheme has zero diffusion (not a particularly good assumption).

Remark 3 *The first term for the numerical diffusion is related to the form of the diffusion operator present in the determination of a cell edge numerical flux. It is formally defined as the difference between the a second-order central difference ap-*

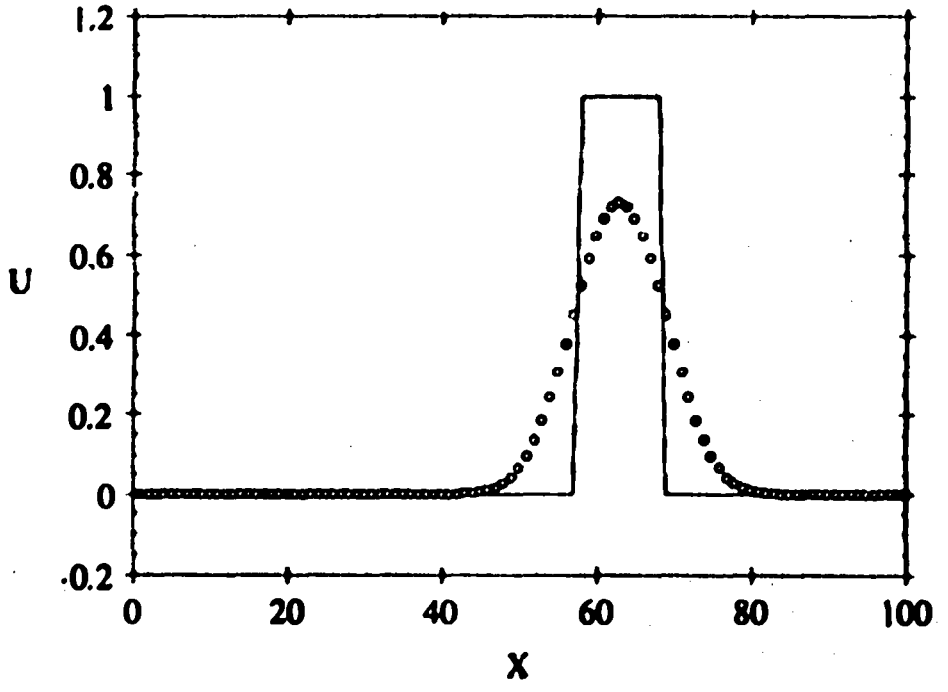


Figure 3.5: The solution for first-order upwind differencing shows the large amount of diffusion present with this algorithm ($a = 1$ and $\nu = 0.5$).

proximation and the numerical flux in a given scheme. The effective induced viscosity is from the numerical error of the scheme and is the coefficient on the second order spatial term.

Remark 4 *Another way to derive this scheme is to assume that each computational cell is interpolated by a piecewise constant profile with the numerical fluxes being based on this reconstruction. Where u is changing, the profile is discontinuous at the cell edges and a solution can be found by solving a local Riemann problem [56]. This is the basic concept in Godunov's method. For the scalar wave equation this results in a scheme identical to the one presented above.*

Figure 3.5 shows the results of using first-order upwinding. The solution's peak is severely clipped and the profile is diffused both in front of and in back of the exact solution. It should also be noted that the solution remains positive definite throughout the computational domain.

3.4 The Lax-Friedrichs Method

The Lax-Friedrichs [55] (sometimes Lax's) method was derived as an answer to the instability of the forward-time centered-space (FTCS) algorithm. It has the following

form,

$$u_j^{n+1} = \frac{1}{2}\left(u_{j-1}^n + u_{j+1}^n\right) - \sigma a\left(u_{j+1}^n - u_{j-1}^n\right) , \tag{3.8}$$

which can be rewritten in conservation form as,

$$f_{j+\frac{1}{2}} = \frac{1}{2}\left[f_{j+1}^n + f_j^n - \frac{1}{\sigma}\left(u_{j+1}^n - u_j^n\right)\right] . \tag{3.9}$$

Looking at the forms of Lax's method and upwinding one can see that the diffusion portion of the flux is always greater than or equal to that found in upwinding. Thus this method has a larger amount of diffusion associated with it than the upwind differenced method. The numerical diffusion is

$$d^{LF} = \frac{\Delta x}{2\sigma} , \tag{3.10a}$$

and again the effective induced viscosity is

$$d^{LF} = \frac{\Delta x}{2\sigma}(1 - \nu) , \tag{3.10b}$$

because this method also produces an exact solution for $\nu = 1$ (see Remark 3).

Figure 3.6 shows the solution obtained with this method, although the solution is positive definite, there are several disturbing features to the solution. One is the terracing of the solution, which gives way to a sawtooth-like structure at the peak of the solution. This is due to the algorithms form which does not require the participation of the information for the j^{th} cell at time step n for the solution of the $n+1$ time step of that cell.

Remark 5 *Interpreted geometrically, the Lax-Friedrichs method is a sort of an "ultra-upwind" method because the solution is over biased (a coefficient grater than one) in the upwind direction. In recent years, the Lax-Friedrichs method has been used with a slight variation. The magnitude of the dissipation in the flux is set to the absolute value of the largest local characteristic speed. For a scalar wave equation, this is identical to the upwind method, but for systems of equations this is much different (this is discussed in more detail in Appendix B).*

3.5 Lax-Wendroff Type Methods

The Lax-Wendroff method [58] is the canonical classical second-order method. This method produces second-order solutions, but with spurious oscillations near discontinuities, thus raising the possibility of producing negative values of positive definite values such as density or pressure. From the standpoint of algorithmic description, geometric depiction is particularly useful. Normally, the method of Lax-Wendroff is

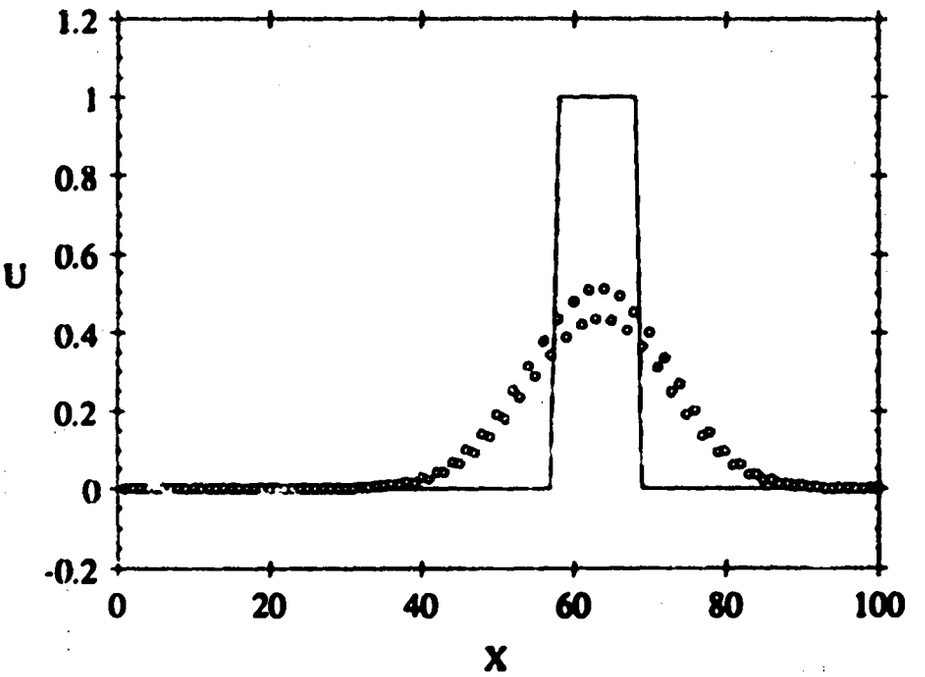


Figure 3.6: The solution for the Lax-Friedrichs method shows the extreme amount of diffusion present with this algorithm. Also noticeable is the terracing and the sawtooth structure in the solution ($a = 1$ and $\nu = 0.5$).

described as a finite-difference algorithm; however, it also can be described geometrically. Figure 3.7 gives a qualitative description of the method.

It is well known that the second-order central difference scheme with forward Euler time differencing is unconditionally unstable. This can be easily verified with VonNeumann stability analysis, but I proceed from a different standpoint. This is motivated by the desire to have a more heuristic explanation for this well-known phenomenon. First, some nomenclature needs to be introduced. The flux functions for difference schemes are functions of the dependent variables and can be written in terms of interpolating polynomials. Thus, given a piecewise polynomial, $P_j(x)$, the flux functions can be written

$$f_j(u) = f[P_j(x)] \; . \tag{3.11}$$

With this definition, the problem reduces to approximating the dependent variables on a grid and computing the value of the interpolant at cell edges.

Returning to the second-order central difference, it can be written as a piecewise

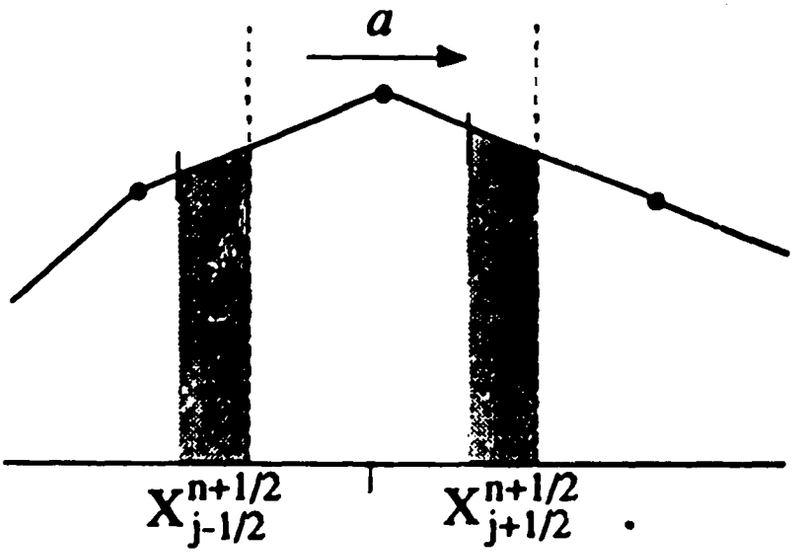


Figure 3.7: The Lax-Wendroff method can be viewed geometrically as a linear interpolation of the initial data with a time centered correction (or time averaged) to the cell edged state. If one thinks of the form of the exact solution to the scalar wave equation, $u(x,t) = u_o(x - a\Delta t)$, this form makes sense.

polynomial on the interval $\left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$ and has the form

$$P_j(x) = \begin{cases} u_j + s_{j+\frac{1}{2}}(x - x_j) \; ; x \in \left[x_j, x_{j+\frac{1}{2}}\right] \\ u_j + s_{j-\frac{1}{2}}(x - x_j) \; ; x \in \left[x_{j-\frac{1}{2}}, x_j\right] \end{cases} , \tag{3.12a}$$

where

$$s_{j-\frac{1}{2}} = \frac{u_j - u_{j-1}}{x_j - x_{j-1}} \text{ , and } s_{j+\frac{1}{2}} = \frac{u_{j+1} - u_j}{x_{j+1} - x_j} \tag{3.12b}$$

This functional form is both C^0 and C^1 continuous. Evaluating the flux function at $x_{j-\frac{1}{2}}$ and $x_{j+\frac{1}{2}}$, the second-order central difference scheme is recovered. This functional form takes absolutely no consideration of the direction of the flow in the problem in finding the numerical flux functions. Perhaps this is a more palatable physically based explanation for the unconditional instability. The method produces spurious oscillations because the solutions computed with these flux functions can lie outside the given values of u.

By considering the fluid motion and in a Lagrangian sense computing the time-centered cell edge positions, which is for the right hand side cell edge

$$x_{r,j} = x_{j+\frac{1}{2}} - \frac{a\Delta t}{2} . \tag{3.13a}$$

and for the left hand side cell edge

$$x_{l,j} = x_{j-\frac{1}{2}} - \frac{a\Delta t}{2} . \tag{3.13b}$$

Inserting these expressions into the second-order central difference polynomials gives the Lax-Wendroff scheme (for a scalar equation). This method is stable for $\lambda a \leq 1$, but still produces spurious oscillations. This stability is solely the result of an "upwind" centered approximation, which now is dependent on the flow direction rather than completely centered in a spatial sense.

Remark 6 *This differs from the account of the Lax-Wendroff method given by LeVeque [40] that requires the direction of the flow to be known in order to define the interpolation.*

The original Lax-Wendroff method [58] uses a second-order accurate Taylor series approximation in time to stabilize the FTCS method. The original derivation was based around the following ideas: given a second-order Taylor series in time

$$u(t + \Delta t) = u(t) + \left.\frac{\partial u}{\partial t}\right|_t + \left.\frac{\partial^2 u}{\partial t^2}\right|_t + \mathcal{O}\left(\Delta t^3\right) , \tag{3.14a}$$

and making substitutions for the time derivatives defines the method. Using the following relations

$$\frac{\partial u}{\partial t} = -\frac{\partial f}{\partial x} . \tag{3.14b}$$

and

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial t}\left(\frac{\partial u}{\partial t}\right) = \frac{\partial}{\partial t}\left(-\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial t}\left(-a\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial x}\left(-a\frac{\partial u}{\partial t}\right) = \frac{\partial}{\partial x}\left(a\frac{\partial f}{\partial x}\right) , \tag{3.14c}$$

gives the final form

$$u(t + \Delta t) = u(t) - \left.\frac{\partial f}{\partial x}\right|_t + \left.\frac{\partial}{\partial x}\left(a\frac{\partial f}{\partial x}\right)\right|_t + \mathcal{O}\left(\Delta t^3\right) , \tag{3.14d}$$

or

$$u(t + \Delta t) = u(t) - \left.\frac{\partial f}{\partial x}\right|_t + \left.\frac{\partial}{\partial x}\left(a^2\frac{\partial u}{\partial x}\right)\right|_t + \mathcal{O}\left(\Delta t^3\right) . \tag{3.14e}$$

The derivatives are all approximated with central differences. The numerical flux functions can be written [110]

$$f_{j+\frac{1}{2}} = \frac{1}{2}\left[(f_j + f_{j+1}) - \sigma a^2 (u_{j+1} - u_j)\right] . \tag{3.15}$$

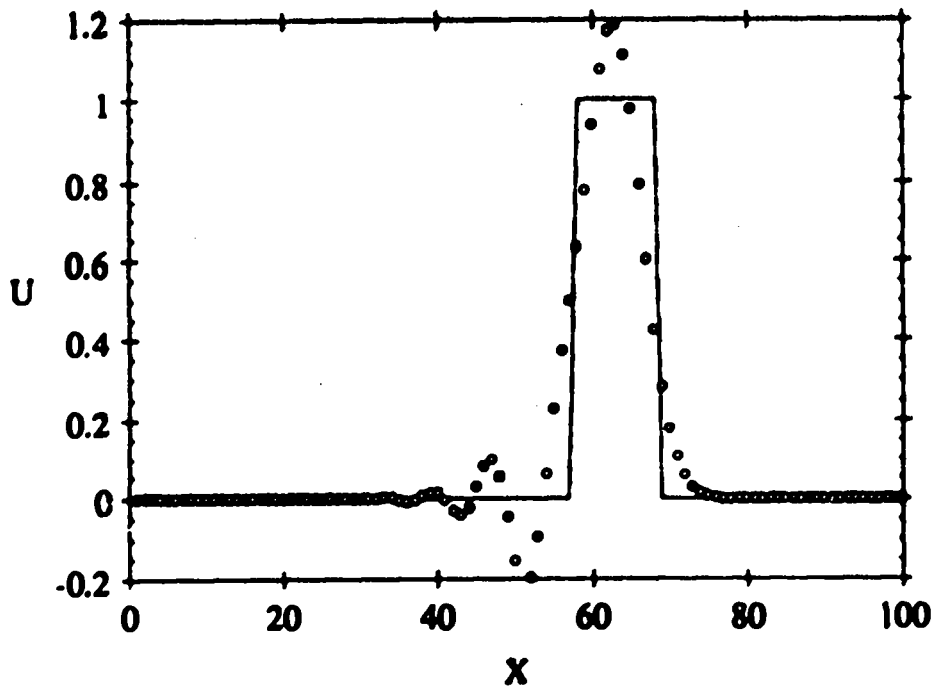


Figure 3.8: Lax-Wendroff's method shows a sharp capture of the discontinuity, but the solution is polluted with dispersive ripples ($a = 1$ and $\nu = 0.5$).

which shows that the numerical diffusion coefficient associated with this method is

$$d^{LW} = \frac{\sigma a^2 \Delta x}{2}, \tag{3.16a}$$

or an effective induced viscosity of

$$d^{LW} = 0. \tag{3.16b}$$

Again, as with the past two methods, the Lax-Wendroff method reproduces the exact solution when used on the scalar wave equation and $\nu = 1$ (see Remark 3). These results do not suggest that this is always possible in the general case; however, they do suggest that the CFL number should be maximized to the extent possible for quality solutions.

The solution found with this algorithm is shown in Fig. 3.8. It shows a sharp location of the discontinuity, but the solution shows a great amount of dispersion and negative values. These values may not be physical as discussed earlier and are aesthetically unappealing. There is also a fairly significant amount of numerical diffusion associated with the fronts. Typically, the Lax-Wendroff method is augmented with artificial diffusion to combat ripples [111, 112].

3.5.1 The Two-Step Lax-Wendroff Method

The Lax-Wendroff method has been reformulated as a two step method, first by Richtmyer [113] and then by Burstein [114]. It can be written as follows,

$$u_{j+\frac{1}{2}}^{n+\frac{1}{2}} = \frac{1}{2}\left(u_j^n + u_{j+1}^n\right) - \frac{1}{2}\sigma a\left(u_{j+1}^n - u_j^n\right) , \tag{3.17a}$$

and a second step

$$u_j^{n+1} = u_j^n - \sigma a\left(u_{j+\frac{1}{2}}^{n+\frac{1}{2}} - u_{j-\frac{1}{2}}^{n+\frac{1}{2}}\right) . \tag{3.17b}$$

This form is already in conservation form. This method is equivalent to the original Lax-Wendroff method for a scalar equation (proven through simple backsubstitution). This formulation has been useful in simplifying the implementation of the Lax-Wendroff method on systems of equations. It may be useful to consider this form (or something similar) in future method development.

3.5.2 MacCormack's Method

MacCormack's method [115] is another derivative of Lax-Wendroff's method and produces similar results. The form of the solution algorithm is as follows,

$$\bar{u}_j = u_j^n - \lambda a\left(u_{j+1}^n - u_j^n\right) , \tag{3.18a}$$

and a second step

$$u_j^{n+1} = \frac{1}{2}\left[u_j^n + \bar{u}_j - \lambda a\left(\bar{u}_j - \bar{u}_{j-1}\right)\right] . \tag{3.18b}$$

In this form, the Lax-Wendroff method appears to be a predictor-corrector method. This method has been particularly important in aerodynamic application where it has found widespread use.

3.6 Second-Order Upwind (Beam-Warming Method)

One classical cure for the problems of the Lax-Wendroff method is to make a second-order scheme with an upwind biased stencil[3]. Using the form (3.12a), this scheme can be defined by setting

$$s_{j+\frac{1}{2}} = \frac{u_j - u_{j-1}}{x_{j+1} - x_j} , \tag{3.19a}$$

if $a > 0$ and

$$s_{j+\frac{1}{2}} = \frac{u_{j+2} - u_{j+1}}{x_{j+1} - x_j} , \tag{3.19b}$$

[3]The term stencil refers to the gridpoints used by a scheme.

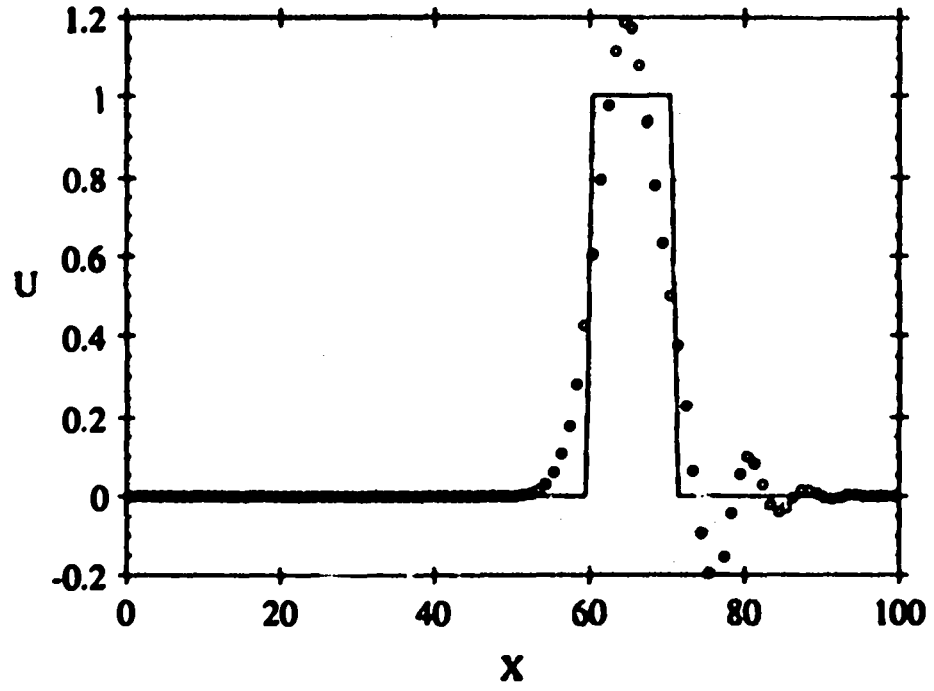


Figure 3.9: The Beam-Warming method shows a sharp capture of the discontinuity, but the solution is polluted with dispersive ripples, but the orientation of the ripples is different than the Lax-Wendroff solution ($a = 1$ and $\nu = 0.5$).

if $a < 0$. With time-centered differencing this is the Beam-Warming scheme [116]. The solution of the test problem is shown in Fig. 3.9.

The methods discussed in this chapter do not cover all "classical" CFD methods, but represent the most commonly used. The concepts presented above also represent the basic means through which modern methods are based. The methods discussed in this chapter are linear. Linearity is expressed in the application of the finite difference stencil to the governing differential equations. In all the classical methods, the stencil is identical for all grid points. The importance of this will become clear shortly.

In the following chapter I describe the basics of high resolution upwind methods for conservation laws. Rather than a fixed finite difference stencil, the methods introduced in the next chapter use adaptive stencils that change as the flow changes. The methods of this chapter are laid as the foundation for what follows.

Chapter 4.

An Introduction to High-Resolution Upwind Shock-Capturing Methods

Linearity breeds contempt. *Peter Lax*

4.1 Motivation

To start the discussion of high-order methods in CFD for solving HCLs, I thought a quick motivational introduction is needed. The first modern method discussed in detail here is that of Godunov [56, 57], which is at the root of most recent methods (see Fig. 2.1). One might believe that using a high quality method like Godunov's would do the job (if more detail is needed, use more grid points). To illustrate why higher order methods are worth exploring, I make use of a test problem used by Woodward and Colella [44]. This is an interacting blast wave problem described in more detail in Appendix A.

In a one-dimensional domain, the density is set to unity everywhere with the fluid at rest, the left most ten percent of domain has pressure set to 1000, the right most 10 percent has a pressure of 100, and the rest of the domain set to 0.01 with $\gamma = 1.4$. Two very strong shocks form and eventually interact forming a combination of shock waves, contact discontinuities and rarefactions. This turns out to be a very stringent test of a numerical method and it is very difficult to resolve all the phenomena involved.

Figures 4.1 and 4.2 show the results for density using Godunov's method (Section 4.3) and a second-order Godunov (Section 4.4) method respectively. The first order Godunov's method uses 5 times the computer memory and 35 times the computer time to solve the problem yet the second-order solution is of much better quality and is closer to the converged solution[1]. This point has been raised in [89], in a simulation of hydrodynamic phenomena in the 1987A supernova. The cost and complexity of the partial parabolic method (PPM) they used allowed the resolution of phenomena in their simulation. With other methods the solutions could not be attempted because of limitations on computer memory. It should be pointed out that as the dimensionality of the problem increases, the advantage of high resolution methods

[1] This is in line with the remarks found in [89]. There it was stated that high resolution second-order (or higher) methods were 15 to 30 times higher in resolution than Godunov's method for contact discontinuities

increases. Things like adaptive griding could also improve matters considerably although a combination of adaptivity and high resolution appears to work best [117].

Remark 7 *The use of 10 times as many grid points implies through the action of the CFL stability criterion that 10 times as many time steps be used for a given calculation. This equals 100 times as many grid points times time steps, which in turn indicates that the high order method is about three times as expensive as Godunov's method on a per grid point per time step basis. From the perspective of performance, at 15 times the resolution the high resolution method is 5 times cheaper per grid point per time step. If these results are applied to multidimensional problems, the differences become more profound.*

4.2 Introduction

The work of Godunov [56] has led to many striking advances in the numerical solution of (2.1). The unique nature of Godunov's work was recognized by van Leer [118]. In a series of papers, he [119, 120, 60] spearheaded the modern development of HOG algorithms. Godunov's method and van Leer's extensions use polynomial representations of the conserved variables in each grid cell in the process of computing the solution. These piecewise polynomials can be discontinuous at grid cell interfaces and as such require some closure relations at these interfaces to compute the numerical fluxes. This closure uses the local solution to a Riemann problem (Appendix B) though either an "exact" [41, 60, 121, 122, 123, 124] or an approximate [125, 126, 63, 127, 128] Riemann solver.

Colella and Woodward [122] advanced the method developed by van Leer in their PPM. This method is still considered a premier method for computing the solutions to (2.1) [129]. Several theoretical advances have been made as well as the more practical ones. Harten's theory of TVD schemes [130, 61] (Section 4.5) made great strides toward understanding the theoretical properties of methods like van Leer's and those discussed below.

Several different varieties of TVD methods have been developed: the modified flux formulation due to Harten and several symmetric TVD schemes. Roe introduced the symmetric form of TVD scheme [131]. Sweby [132] and Davis [133] also presented methods of the same general form. These were all derived as a Lax-Wendroff method augmented with a nonlinear upwind biased dissipation term. Yee [134] christened these schemes as symmetric TVD schemes in her paper. The general form of symmetric TVD schemes can be viewed in different ways: as an advanced form of artificial diffusion, and as a Lax-Wendroff [58] with an additional dissipative flux to ensure a TVD solution. Along other lines, Goodman and LeVesque [135] took a geometric view similar to van Leer's work in deriving a TVD method.

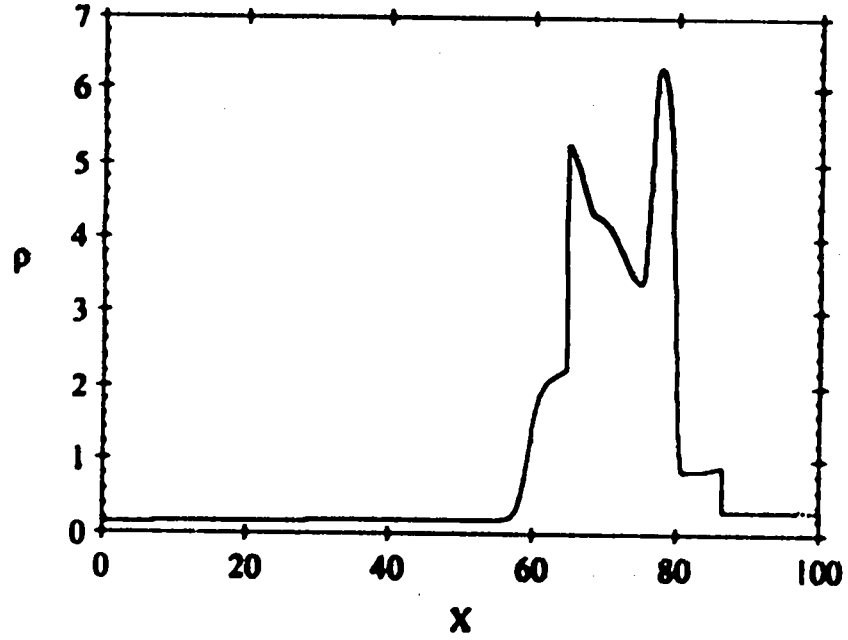


Figure 4.1: The density computed with Godunov's method using 10,000 grid points shows the general structure of the solution; however, the solution also shows significant smearing behind the contact discontinuity at $x \approx 0.6$. The peaks at $x \approx 0.65$ and $x \approx 0.80$ are clipped. ($\Delta x = 0.01, \nu = 0.99, t = 3.80$.)

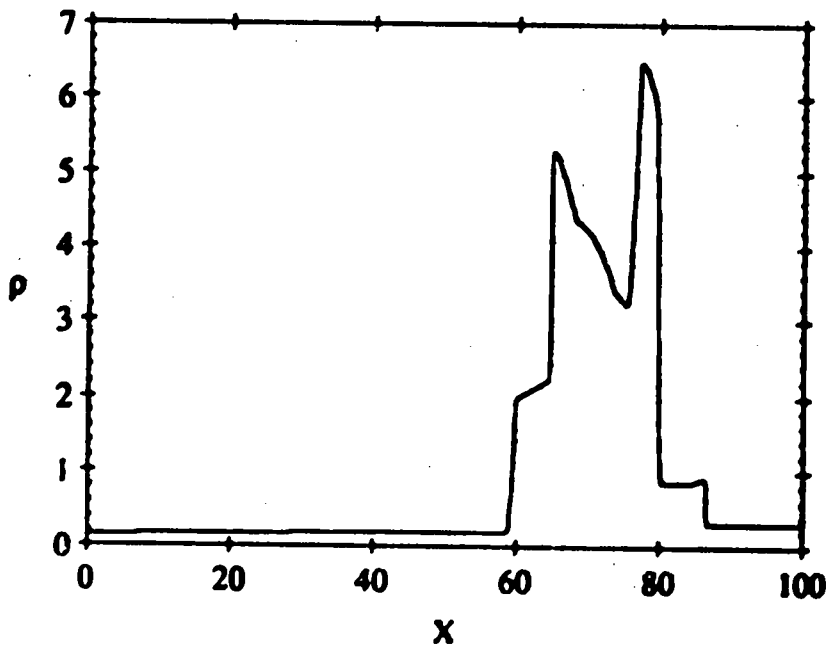


Figure 4.2: The density computed with a second-order Godunov method using 1000 grid points shows a nearly converged solution. Much of the smearing and clipping present in the first-order solution is gone. (See Woodward and Colella 1984 for the converged solution.)

A limitation of these methods is that they are limited to first-order accuracy in the maximum error norm. This is due to the action of the flux or slope limiter used in assuring the TVD quality of the solutions. To increase the accuracy of this sort of method, more elaborate numerical algorithms have been developed in the past few years. Among these are the uniformly non-oscillatory (UNO) scheme of Harten and Osher [136], which is second- order accurate in all norms. Essentially non-oscillatory (ENO) methods are described in a series of papers [64, 137, 65, 66], where these ideas have been extended to arbitrarily high orders of accuracy (Section 4.4.2). These ideas are also making their way into multidimensional algorithms [138, 139].

Another modern advection algorithm also can be viewed along these lines. Perhaps the first modern algorithm to recognize the necessity of nonlinearity in the difference scheme was the FCT method as introduced by Boris and Book [59] (Section 4.6). This method was developed with the recognition of the theorem of Godunov,

Theorem 3 (Godunov [56]) *No monotone numerical algorithm for solving (2.1) can be both linear and second-order accurate.*

This does not preclude the possibility of producing a "monotone" second-order scheme, but simply state that such a method cannot be linear in nature. Thus the FCT is designed as a nonlinear blending of high- and low-order numerical fluxes, which ensures the lack of dispersive ripples. In a series of papers [59, 140, 141, 142, 62] this method has been revised and extended.

Digressing slightly, there appears to be a schism in the literature between the TVD, HOG and ENO type methods and the FCT methods. Authors doing research on each method usually mention the other methods, but the synergism ends there. It is often stated in the FCT literature that the TVD type methods require Riemann solvers and as such are horrendously complex in comparison to FCT. It is my contention that this is simply not true. Underlying each method is a scheme for scalar advection, which is at the genesis of more complex development. In extending the methods to systems of equations, the TVD type methods use Riemann solvers, which have many exceptional theoretical and aesthetic appeals. The extension of FCT, on-the-other-hand, is usually extended in what seems an *ad hoc* or *naive* (see Section B.3.4) formulation [143, 144].

Borrowing from [45] one can sort of "see" how various schemes are related pictorially. This is done in Fig. 4.3. If one imagines some sort of space of schemes with monotone schemes, S_M being the most restrictive and the space of all transport schemes, S_T encompassing all methods. The various methods can be seen as a set of overlapping spaces. The space of all TVD methods is $S_{TVD} \cup S_M$ and ENO schemes are the union of the TVD space and that labeled S_{ENO}.

Recently, I have thought a lot about the philosophy related to the design of high resolution schemes and I believe these philosophies can be classified as follows:

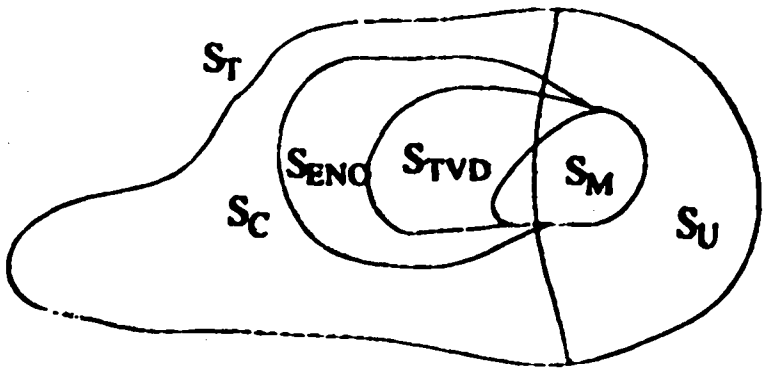


Figure 1.3: In this diagram a rough classification of modern numerical schemes is shown. S_U is the space of upwind methods and S_C is the space of centered schemes, the other terms are explained in the text. (adapted from [45, 145].)

1. Artificial Viscosity: There are those that believe that the high-order schemes are simply fancy artificial diffusion prescriptions. This is largely a product of the TVD-Lax-Wendroff [133, 131, 132] and the symmetric TVD [134] methods.

2. Hybridization: The FCT [59, 140, 141, 142, 62] and Hybrid [146] methods are most easily classified as combinations of first- and higher-order classic schemes.

3. Mathematical Theory: Harten [130, 61] and Harten et. al [64] have produced a mathematical framework which is useful in producing rigorous proofs and bounds on the behavior of these schemes (TVD) and a vague generalization to less restrictive schemes (ENO).

4. Interpolation and Advection: This was given by van Leer [120, 147] (based on the work of Godunov) and then extended in PPM. The method seems somewhat heuristic in nature, although it works well. TVD theory aids and expands this train of thought, which works well for conceptualization of the schemes. The ENO algorithms extend this view to a broader class of methods, but at this point do not include the breadth of possible methods. In a recent paper, Harten brings the arguments of semi-Lagrangian method [112] into the arena of high-resolution methods. This should be clarified by the fact that unlike those methods used in meteorological [148, 149] flow by $\nu \leq 1$. Despite this kind of different viewpoint, the results are generally similar, although the meteorological schemes are not conservative in nature. Thus they are not appealing for computations of discontinuous solutions.

At some point, these various approaches should be equivalent, which would result in an increased synergism between methods and ease of analysis.

Remark 8 *In [149] it was noted that van Leer began looking at semi-Lagrangian methods early in his studies, but dropped them from consideration because of their*

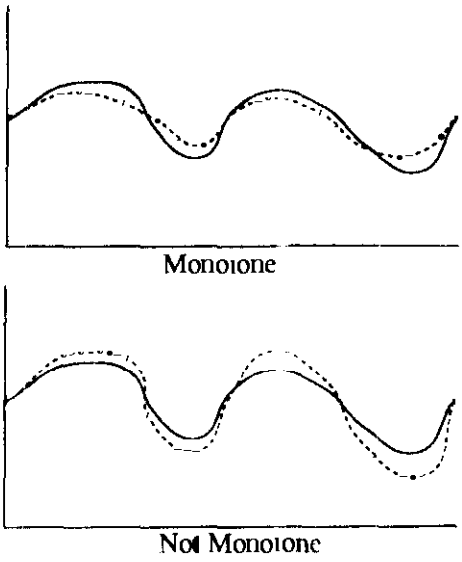


Figure 4.4: The initial data is denoted by the solid line while the dotted line shows the solution at some advanced time on a periodic domain. The upper figure's solution is monotone because the extrema in the advanced time solution are bounded above and below by the initial data. The lower figure's solution is not monotone because new extrema exist in the solution.

lack of conservation.

A key concept in this entire discussion is that of monotone convection[2]. This means that the solution is a physical solution for physical initial data and that it does not create new extrema in the solution. This is depicted graphically in Fig. 4.4.

Definition 1 (Monotone Numerical Advection [151]) *Monotone numerical advection is defined by a scheme which is a combination of coefficients of the local data which are all positive. Consistency requires that some conservation principle be enforced i.e. the coefficients sum to one. This also means that the numerical scheme does not introduce new extrema into the solution.*

For the remainder of the presentation, the following nomenclature is used: $\Delta_{j+\frac{1}{2}} u = u_{j+1} - u_j$. A conservative finite-difference solution to (2.1) using a simple forward Euler time discretization is

$$u_j^{n+1} = u_j^n - \sigma \left(f^n_{j+\frac{1}{2}} - f^n_{j-\frac{1}{2}} \right) . \tag{4.1}$$

[2]Defined rigorously, monotone convection implies that the finite difference scheme is first order [73]. Also some work shows that as currently defined no scheme can be TVD in more than one dimension [150].

The temporal spacing is Δt and Δx is the spatial mesh spacing. The superscript n refers to time, t, $n+1$ refers to the time $t+\Delta t$, and the subscript j refers to space with j being a cell center and $j \pm \frac{1}{2}$ being the cell edges. The construction of the numerical fluxes $\dot{f}_{j\pm\frac{1}{2}}$ is at the heart of this subject. The cell edge flux can be defined as

$$\dot{f}_{j+\frac{1}{2}} = \frac{1}{2}(f_j + f_{j+1}) + \phi_{j+\frac{1}{2}} , \tag{4.2a}$$

where ϕ is a numerical dissipation term. For a system of equations the flux is written

$$\dot{\mathbf{F}}_{j+\frac{1}{2}} = \frac{1}{2}(\mathbf{F}_j + \mathbf{F}_{j+1}) + \Phi_{j+\frac{1}{2}} , \tag{4.2b}$$

where $\mathbf{F}$ and Φ are vectors, but are defined similarly to the single equation case. For instance, the first-order donor-cell flux can be written

$$\dot{f}^{DC}_{j+\frac{1}{2}} = \frac{1}{2}\left(f_j + f_{j+1} - \left|a_{j+\frac{1}{2}}\right| \Delta_{j+\frac{1}{2}} u\right) , \tag{4.3}$$

thus

$$\phi^{DC}_{j+\frac{1}{2}} = -\frac{1}{2}\left|a_{j+\frac{1}{2}}\right| \Delta_{j+\frac{1}{2}} u .$$

Remark 9 *When numerical schemes become nonlinear in nature and/or are applied to nonlinear problems, standard means of analysis are not typically valid. New approaches to method analysis have been developed, but are not as mature as classical methods. LeVeque [40] gives an overview of this topic. Much of the modern analysis is based on "compensated compactness" as used by DiPerna [152, 153] in his proofs of convergence. Nonlinear dynamics may also yield useful means of analysis [154].*

4.3 Godunov's Method

I have already visited Godunov's method in the Section 3.3. For a single scalar equation this is simply the upwind method described there. For nonlinear problems this is not so straightforward. The key point in constructing a Godunov method is to use some sort of Riemann solver. Another consideration is entropy satisfaction of the solution [155]. This generally means that the solution must contain sufficient numerical viscosity to insure physical solutions.

The following algorithm gives a general outline for Godunov type methods.

Algorithm 1 *[Godunov's Method]*

1. **(Initialization Step)** Average the initial distribution over the computational cells

$$u^0_j = \frac{1}{\Delta x} \int_{x_j - \Delta x/2}^{x_j + \Delta x/2} u(x)\, dx . \tag{4.4a}$$

2. (Reconstruction step) Reconstruct the initial distribution as piecewise polynomials over the computational cells

$$u_j(x) = P_j(x) \ , \tag{4.4b}$$

where $P_j(x)$, $x \in [x_j - \Delta x_j/2, x_j + \Delta x_j/2]$ is a polynomial in cell j.

3. (Solution in the Small Step) Solve the initial value problem at each cell interface where discontinuities can exist

$$u(x,t) = E(x, t - t^n) \cdot u(x, t^n) \ , \tag{4.4c}$$

where $E(x, t - t^n)$ symbolically represents the evolution operator given by the solution to the Riemann problem.

4. (Averaging Step) Reaverage the solution over the grid cells given the solution operator in the previous step.

$$u_j^{n+1} = \frac{1}{\Delta_j x} \int_{x_j - \Delta_j x/2}^{x_j + \Delta_j x/2} u\left(x, t^{n+1}\right) dx \ . \tag{4.4d}$$

5. Go back to the reconstruction step.

This process is shown schematically in Fig. 4.5.

Remark 10 *Osher [155] defined a Godunov flux for scalar equations as*

$$\begin{aligned} &(a)\ \text{if } u_j < u_{j+1} \text{ then } f^G_{j+\frac{1}{2}} = \min(u), u \in [u_j, u_{j+1}] \\ &(b)\ \text{if } u_j > u_{j+1} \text{ then } f^G_{j+\frac{1}{2}} = \max(u), u \in [u_j, u_{j+1}] \end{aligned} \ , \tag{4.5a}$$

and the inequality for an entropy satisfying flux is

$$\begin{aligned} &(a)\ \text{if } u_j < u_{j+1} \text{ then } f_{j+\frac{1}{2}} \le f^G_{j+\frac{1}{2}} \\ &(b)\ \text{if } u_j > u_{j+1} \text{ then } f_{j+\frac{1}{2}} \ge f^G_{j+\frac{1}{2}} \end{aligned} \ . \tag{4.5b}$$

For the scalar equation, this Godunov flux is the least diffusive entropy satisfying flux. Thus for the case of scalar equations one can show what the appropriate entropy inequalities are. This inequality can be written

$$sign(u_{j+1} - u_j)\left[f_{j+\frac{1}{2}} - f(u)\right] \le 0 \ , u \in [u_j, u_{j+1}] \ . \tag{4.5c}$$

Osher defined schemes which meet the entropy requirements as "E-schemes". This concept has proven to be important in the development of higher order schemes which

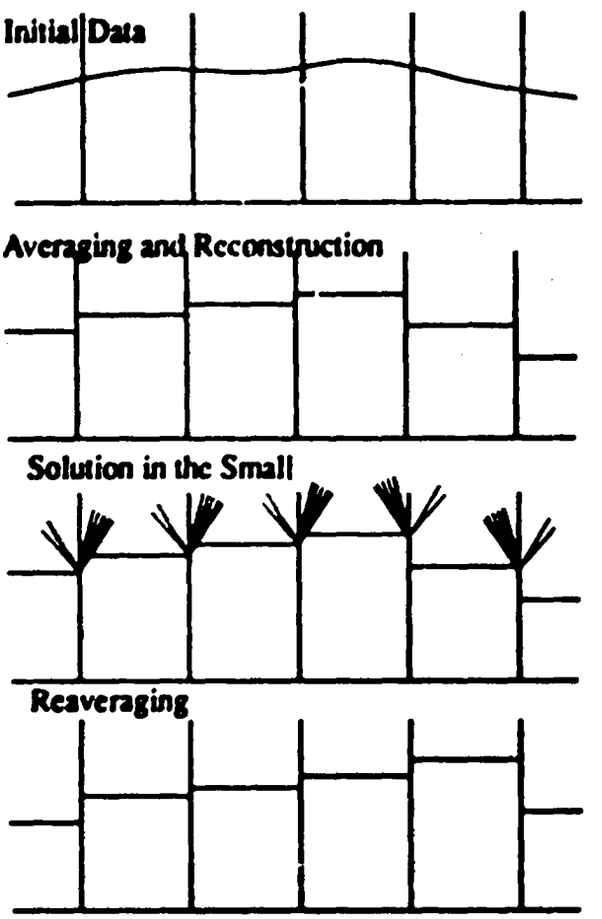


Figure 4.5: The following steps are shown: averaging and reconstruction, solution in the small, and reaveraging in this schematic representation of Godunov's method

produce physical solutions. It is common practice to develop the higher order schemes with an E-scheme as a building block.

This algorithm can be formulated in several ways: in a fixed or Eulerian coordinate system or in a moving or Lagrangian coordinate system. With the Lagrangian formulation, the common practice is to set the coordinate frames speed equal to that of the flow. Another common practice is to compute solutions in the Lagrangian frame and map the results back to an Eulerian grid. For the Eulerian algorithm, the solution in the small is done in a fixed coordinate frame so the averaging step is a simple one step process. In the Lagrangian algorithm, the averaging step takes place in two steps: first an average in the Lagrangian frame and then a remap to the fixed Eulerian grid.

The averaging step can be simplified with the divergence theorem that allows the integral

$$u_j^{n+1} = \frac{1}{\Delta x_j} \int_{x_j - \Delta x_j/2}^{x_j + \Delta x_j/2} u\left(x, t^{n+1}\right) dx \,,$$

to be transformed to

$$u_j^{n+1} = u_j^n - \lambda \left(f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}} \right) \,, \tag{4.6}$$

where $\lambda = \Delta t / \Delta x$ and

$$f_{j+\frac{1}{2}} = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f\left(x_{j+\frac{1}{2}}, t\right) dt \,. \tag{4.7}$$

This formulation is just like the normal finite difference equations for a differential equation in conservation form. For the solution in Lagrangian coordinates, the spatial variable x in the above equations is replaced with ξ, the mass variable. The remap step of the Lagrangian Godunov also can be expressed in these terms. In this step, the solution in the Lagrangian coordinates is mapped onto an Eulerian grid. This can be expressed as the advection of the conserved quantities through the cell boundaries.

This reaveraging step (see Appendix B equations (B.3a)-(B.3c)) can be derived from the concept of operator splitting [156]. The Lagrangian step is the solution for the Euler equations for the sound wave related transport and the remap is the solution for the advection related transport. This concept is at the genesis of the Arbitrary Lagrangian-Eulerian algorithms [157], but these differences are more philosophical than substantive.

The remapping procedure must deal with several specific possibilities, as shown in Fig. 4.6. Carrying out the summations over the Eulerian grid cells reveals that the use of a simple upwind difference formula suffices to carry out the remapping. From the solution of the Lagrangian equations the cell edge velocities are known, thus the

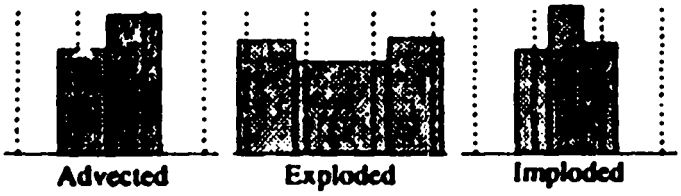


Figure 4.6: The cases which must be considered by a remap algorithm.

remapping is uniquely determined. The formula is constructed as follows

$$\phi_j^{n+1} = \frac{\Delta \tilde{x}}{\Delta x}\phi_j - \frac{\Delta t}{\Delta x}\left[\hat{f}\left(\phi_{j+\frac{1}{2}}\right) - \hat{f}\left(\phi_{j-\frac{1}{2}}\right)\right] , \tag{4.8a}$$

with

$$\hat{f}\left(\phi_{j+\frac{1}{2}}\right) = \frac{1}{2}\dot{u}_{j+\frac{1}{2}}\left(\phi_j + \phi_{j+1}\right) - \frac{1}{2}\left|\tilde{u}_{j+\frac{1}{2}}\right|\left(\phi_{j+1} - \phi_j\right) , \tag{4.8b}$$

where all quantities with a "tilde" are new time Lagrangian frame variables except $\dot{u}$, which is time centered.

The formulation above has several stability limits. For the solution step to make sense [30] requires that the waves not interact which leads to the restriction

$$\Delta t \leq \inf_j \left(\frac{\Delta x_j}{2|a_j|}\right) , \tag{4.9}$$

where a_j is the maximum wavespeed present in each cell. This means that waves cannot pass through more than half a grid cell in a time step. The stability restriction is the more familiar Courant-Friedrichs-Lewy (CFL) condition

$$\Delta t \leq \inf_j \left(\frac{\Delta x_j}{|a_j|}\right) , \tag{4.10}$$

which is the restriction usually taken for methods of this type. For the purely Eulerian calculations with the Euler equations, see (B.1a)-(B.1c),

$$\Delta t \leq \inf_j \left(\frac{\Delta x_j}{|u_j - c_j|}, \frac{\Delta x_j}{|u_j + c_j|}\right) .$$

where c_j is the Eulerian sound speed. For the Lagrangian computations with the remap step, see (B.2a)-(B.3c), there are three restrictions to consider:

$$\Delta t \leq \inf_j \left(\frac{\Delta \xi_j}{C_j}, \frac{\Delta x_j}{|u_j|}, \frac{\Delta x_j}{\Delta_j u}\right) ,$$

where $C_j = \rho c_j$ the Lagrangian sound speed and the sound speed restriction refers to the Lagrangian step, the advective velocity is for the remap step, and the zone

tangling limit.

4.4 High-Order Godunov Methods

For Godunov's method, the reconstruction step consists of setting

$$P_j(x) = u_j^n \ .$$

or piecewise constant. The Eulerian Godunov uses the Eulerian equation set for the solution step, while the Lagrangian with remap Godunov uses the Lagrangian equations with an averaging done in the moving coordinate frame followed by the remap step back to the Eulerian grid (see Appendix B).

Remark 11 *The primary (and often the only) difference between Godunov's method, which is first order accurate, and higher order methods (see Section 4.4) like MUSCL [60] and PPM [122] is the order of the polynomial used in the reconstruction step.*

Further developments on this topic were achieved by van Leer [60] in his higher order extensions of Godunov's method often referred to as monotone upstream-centered scheme for conservation laws (MUSCL). Recently, researchers have extended the ideas of van Leer to arbitrarily high-order spatially or temporally and christened these methods as uniformly [136] or essentially [64] non-oscillatory (UNO or ENO) schemes.

4.4.1 MUSCL Type Schemes

The second-order methods developed by van Leer essentially replaced the constant piecewise profile used in Godunov's method with a linear profile. This profile is "limited" (Section 4.7 and Chapter 8) in order to prevent non-monotone behavior in the solution procedure. Van Leer's criteria was somewhat heuristic in nature, although it turns out to be fairly rigorous after Harten's work on the theory of TVD schemes [130, 61]. The criteria states that the interpolation in a given cell should not lie outside the range of values defined by the cell average and the neighboring values of the variable being interpolated [120, 27]. This is shown in Fig. 4.7. Stated mathematically this is

$$\min_{|k-j|\le 1} u_k^n \le P_j(x) \le \max_{|k-j|\le 1} u_k^n \ . \tag{4.11}$$

Woodward states that this can be relaxed slightly to the averages of the advected quantity within a cell and that which remains in its original cell must lie within the range of the original cell average and its neighbors. A scheme typical of those used here is

$$P_j(x) = u_j + \widetilde{\Delta_j u}\ \frac{x - x_j)}{\Delta_j x} \ . \tag{4.12}$$

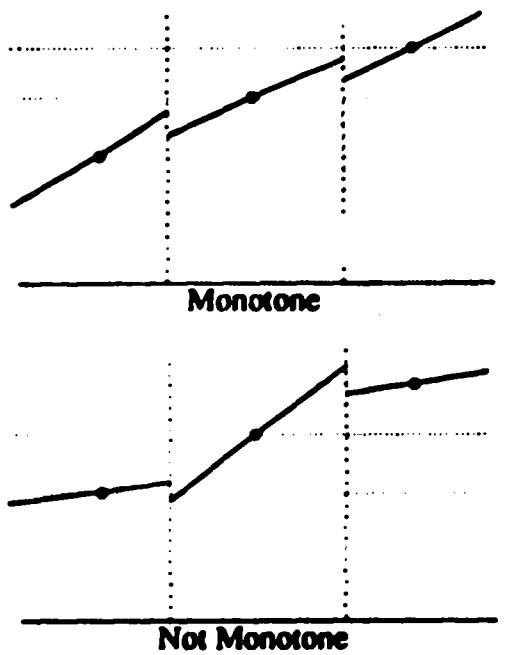


Figure 4.7: A graphical depiction of van Leer's heuristic monotonicity constraint. For the second constraint given by Woodward the interpolation is monotone for some time step sizes.

where $\widetilde{\Delta_j u}$ is a limited approximation to $du/dx|_j \Delta_j x$.

With a second-order algorithm, the question of time accuracy must be addressed. This is usually done through a Lax-Wendroff like procedure like that described in the previous chapter. This can proceed from two viewpoints: the first being that I am moving with the fluid to the point in time which is the average of the old and new time steps and evaluating the polynomial reconstruction there, the second is that of averaging the polynomial over the domain of dependence for the time step [122]. These two views are equivalent if the integral time average is evaluated with a midpoint rule. This process is depicted in Fig. 4.8.

Van Leer [158, 159] reports another approach to finding a second-order accurate temporal solution. Defining $u_{j,l}$ as the value at the left cell edge of cell j and $u_{j,r}$ as the value at the right hand cell edge of j, the second-order time accurate values of $u_{j,l}$ and $u_{j,r}$ are computed from

$$u_{j,l}^{n+\frac{1}{2}} = u_{j,l}^{n} - \frac{\sigma}{2}\left[f\left(u_{j,r}^{n}\right) - f\left(u_{j,l}^{n}\right)\right] \tag{4.13a}$$

and

$$u_{j,r}^{n+\frac{1}{2}} = u_{j,r}^{n} - \frac{\sigma}{2}\left[f\left(u_{j,r}^{n}\right) - f\left(u_{j,l}^{n}\right)\right] . \tag{4.13b}$$

This form of the algorithm bears great resemblance to the two-step Lax-Wendroff scheme presented in Section 3.5.1. Similar sorts of ideas are also expressed in a series

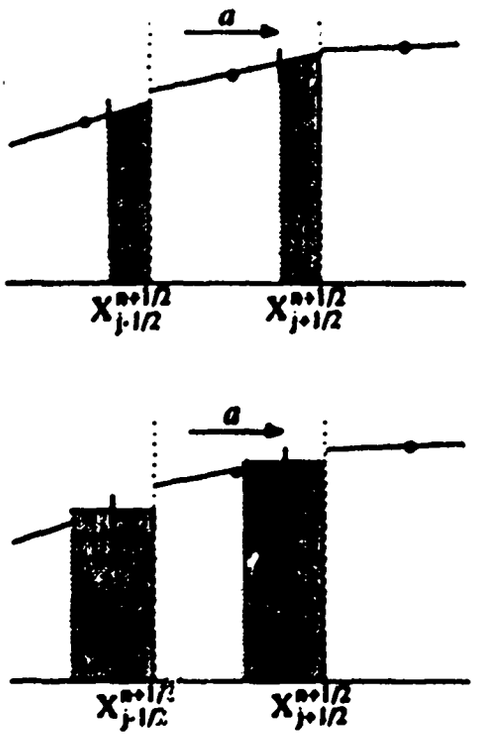


Figure 4.8: Two views of time accurate computation of cell edge values.

of papers [65, 160, 66] where a TVD Runge-Kutta time discretization is introduced and implemented.

Remark 12 *The TVD Runge-Kutta temporal discretization provides the means through which high-order temporal accuracy can be achieved without significant implementation difficulties. This is especially true in multidimensional problems or with systems of HCLs. These multistage algorithms can be written in the following form*

$$u^i = \sum_{k=0}^{i-1} \left[\alpha_{ik} u^k + \beta_{ik} \Delta t L\left(u^k\right) \right] , \tag{4.14a}$$

where the discrete differential operator is denoted by

$$\frac{\partial u}{\partial t} = L(u) , \tag{4.14b}$$

and α_{ik} and β_{ik} are coefficients. The criteria for this to produce TVD results (see Section 4.5) given an appropriate spatial operator is a CFL condition

$$\nu \leq \frac{\alpha_{ik}}{|\beta_{ik}|} . \tag{4.14c}$$

If β_{ik} is negative, the spatial operator must be antiupwind [65, 160]. In those references

a number of schemes are defined.

4.4.2 ENO Type Schemes

Harten and Osher [136] defined a new class of schemes as being uniformly non-oscillatory. This class of method is part of and predecessor to the ENO schemes. One particularly distinguishing fact about this scheme is that it is second-order accurate in all its norms. This gives it some strong advantages over other second-order high resolution schemes, which degenerate to first-order accuracy in the maximum norm.

Definition 2 (Harten and Osher [136]) *Non-oscillatory interpolation is defined by interpolation, $P_j(x)$ that has its number of extrema in an interval that is not exceeded by the local extrema in the data, $u(x)$.*

The construction of ENO schemes has extended the concept of high-order Godunov methods to a much wider range of potential schemes [161] (this class of methods included other Godunov type algorithms). The basic concept of the ENO schemes is to compute a interpolating polynomial using the data from the smoothest part of the grid locally [162]. To do this a limiter is used to choose which direction to go for the smoothest reconstruction. Thus the stencil used for the finite difference formulas is adaptive in nature and the accuracy of the scheme is limited only by its implementation and the properties of the data. One problem is that despite the relatively simple concept, the ENO schemes [64] as originally formulated are horribly complex. This problem is even more severe in multi-dimensional implementations [161, 64]. Shu and Osher [65, 66] have eased this burden somewhat and if more recent work is any indication [139] this should ease more. For ENO schemes, in general, most properties such as convergence, boundedness of solutions etc. have yet to be proven.

Definition 3 (Harten, Osher, Engquist and Chakravarthy [64]) *Essentially non-oscillatory interpolation is defined by interpolation, $P_j(x)$ that is the smoothest approximation to the data in some sense.*

An ENO algorithm for polynomial reconstruction is outlined below. This is known a reconstruction by a primitive function. This ENO formulation is based on the interpolation of a function defined by

$$Q\left(x_{j+\frac{1}{2}}\right) = \int_{-\infty}^{x_{j+\frac{1}{2}}} u\, dx\,, \tag{4.15a}$$

thus

$$u_j(x) = \frac{dQ_j(x)}{dx}\,. \tag{4.15b}$$

By virtue of the previous two equations, the interpolation can be integrated to the cell average of cell j, but also every cell the stencil for cell j.

Before showing the algorithm, some terms need to be defined

$$a^k = Q\left[x^{k-1}_{j,min}, \ldots, x^{k-1}_{j,max+1}\right] , \tag{4.16a}$$

$$b^k = Q\left[x^{k-1}_{j,min-1}, \ldots, x^{k-1}_{j,max}\right] , \tag{4.16b}$$

where the brackets denote the k^{th} divided difference [163] which can be defined recursively[3] The algorithm computes a polynomial for $Q\left(x_{j+\frac{1}{2}}\right)$, which once differentiated can serve as the polynomial approximation in the j^{th} cell.

Algorithm 2 *[ENO Reconstruction via Primitive Function [64]]*

1. Initialize $k = 0$, $x^0_{j,min} = x^0_{j,max} = x_{j+\frac{1}{2}}$
2. If $\left|a^k\right| \geq \left|b^k\right|$ then
$$c^k = b^k , \tag{4.17a}$$
$$(j,min)^{k+1} = (j,min)^k - 1 , \ (j,max)^{k+1} = (j,max)^k . \tag{4.17b}$$
3. If $\left|a^k\right| < \left|b^k\right|$ then
$$c^k = a^k , \tag{4.17c}$$
$$(j,min)^{k+1} = (j,min)^k , \ (j,max)^{k+1} = (j,max)^k + 1 . \tag{4.17d}$$
4. $k = k + 1$
5. Return to step 2 until desired accuracy is achieved (k=n).
6. Define the following polynomial
$$P(x) = \sum_{k=1}^{n} c^k \prod_{i=j,min^k}^{j,max^k} (x - x_i) . \tag{4.17e}$$

Remark 13 *The consideration of point values versus cell averages is of paramount importance in a theoretical sense. Godunov's method is predicated on the concept that the grid point values are averages over a control volume. The spatial determination of the values is only set in the averaged sense, but the point values are not defined clearly as to where they should reside in space. This is a sort of grid uncertainty problem or Gibb's error. Because most ENO implementations are based on interpolating*

[3] A divided difference is defined as $Q[x_1, \ldots, x_n] = (Q[x_2, \ldots, x_n] - Q[x_1, \ldots, x_{n-1}]) / (x_n - x_1)$.

$Q(x)$ this problem does not arise. From the standpoint of conservation the interpolation methodology is not crucial. It is precisely this point on which the problem of implementation of ENO schemes hinges. See Chapter 9 for further discussion of this topic.

Remark 14 *In Shu and Osher's papers on the easy implementation of ENO schemes, a formula was presented without much explanation. Their numerical flux is defined by*

$$\hat{f}_{j+\frac{1}{2}} = f_{j+\frac{1}{2}} + \sum_{k=1}^{m} a_{2k} \left(\frac{\partial^{2k} f}{\partial x^{2k}} \right)_{j+\frac{1}{2}} + \mathcal{O}h^{m+1} ,$$

where $a_2 = \frac{-1}{24}$ and $a_4 = \frac{7}{5760}$. Where does this come from? From earlier ENO work

$$\hat{f}_{j+\frac{1}{2}} = \left. \frac{dQ}{dx} \right|_{j+\frac{1}{2}} ,$$

where

$$Q_{j+\frac{1}{2}} = \int_{x_o}^{x_{j+\frac{1}{2}}} f(x)\, dx .$$

From Hildebrand's numerical analysis text [163], the coefficients in the above equation are from the Euler-MacClaurin equation for errors in integration with a slight modification to take the function to approximate $\hat{f}_{j+\frac{1}{2}}$ rather than $\hat{f}_{j-\frac{1}{2}}$ as the equation in the text would indicate. This corresponds to adding $\hat{f}_{j+\frac{1}{2}}$ to the equation and reversing the signs of the error terms. This raises the question of whether or not the Q function is correct in the sense that this implies. The definition of the point values as cell averages would support this, but it raises questions of the correct derivation of these concepts in multidimensions especially on non-orthogonal grids or unstructured grids [16].

To close out this section, the results on the same test problem used for the classical methods is used with a high-order Godunov method. The results shown in Fig. 4.9 are much better than those found by any of the classical method, with the discontinuities remaining sharp and with little smearing and no creation of oscillations.

Remark 15 *One problem with this sort of method is that it is expensive to use in some cases. Some promising work has appeared recently which only applied to more complex methods described above at a few grid locations (where oscillations would occur with classical methods). These method use a filtering technique to choose where to apply the HOG-type methods [164, 165].*

4.5 Total Variation Diminishing Methods

The effort to put the new modern algorithms on firmer theoretical footing resulted in the concept of total variation diminishing (TVD) methods [130], which have a number

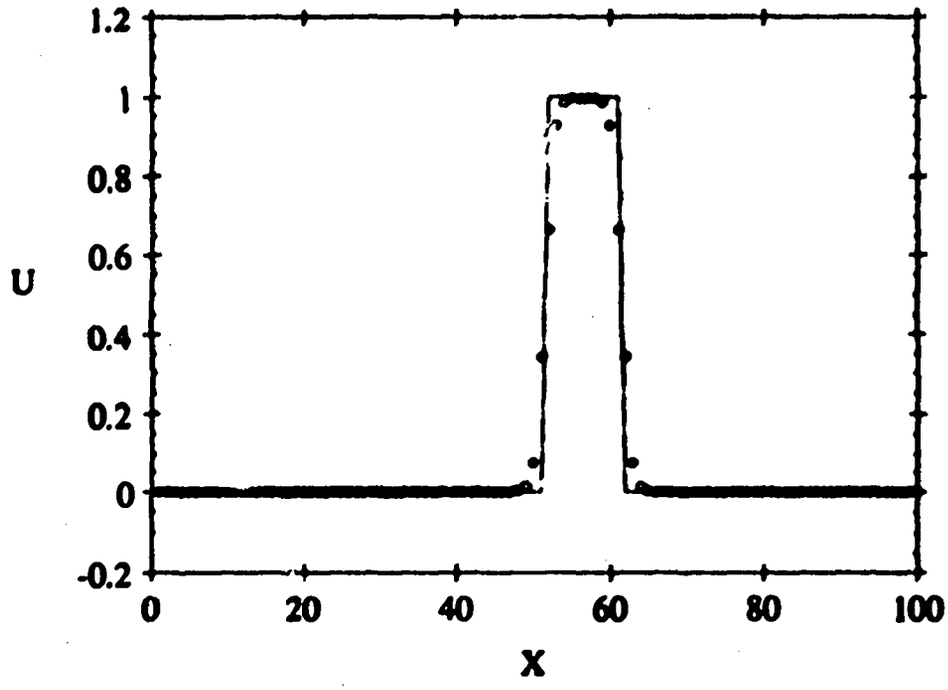


Figure 4.9: Computation of a square wave by the scalar wave equation using a HOG algorithm ($a = 1.$ and $\nu = 0.5$).

of desirable properties. To be total variation diminishing, a scheme must satisfy the following inequalities,

$$TV\left(u^{n+1}\right) \leq TV\left(u^{n}\right) ,$$

where

$$TV(u) = \sum_{j=-\infty}^{\infty} |u_{j+1} - u_j| .$$

While these methods include classic monotone schemes (such as upwind differencing or Lax-Friedrichs), they can also be extended to include methods that are second-order in the L_1 norm. By construction, these methods are still first-order at points of extrema (in the L_∞ norm). A second property of TVD schemes, which is both useful and satisfying, is that they can be extended to include implicit temporal differencing [110]. This generality is quite desirable as it allows a more general use of TVD algorithms for a wide range of problems and applications. It should be noted that MUSCL schemes have also been extended to include implicit temporal differencing.

The basic proof of the TVD property proceeds as follows:

Theorem 4 (Harten [130]) *Given a scalar wave equation and a conservative numerical scheme written as*

$$u_j^{n+1} = u_j^n + C^+_{j+\frac{1}{2}} \Delta_{j+\frac{1}{2}} u^n - C^-_{j-\frac{1}{2}} \Delta_{j-\frac{1}{2}} u^n , \qquad (4.18a)$$

where

$$C^-_{j+\frac{1}{2}} \geq 0\,,\; C^+_{j+\frac{1}{2}} \geq 0\,, \tag{4.18b}$$

and

$$C^-_{j+\frac{1}{2}} + C^+_{j+\frac{1}{2}} \leq 1\,, \tag{4.18c}$$

then the scheme is TVD.

Proof. Start by subtracting the equations at $j+1$ from j giving

$$\Delta_{j+\frac{1}{2}}u = C^-_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u + \left(1 - C^-_{j+\frac{1}{2}} - C^+_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u + C^+_{j+\frac{1}{2}}\Delta_{j+\frac{1}{2}}u\,. \tag{4.19a}$$

Because I am assuming the condition stated in the theorem, all the terms on the right hand side are positive, thus by the triangle inequality

$$\left|\Delta_{j+\frac{1}{2}}u\right| \leq C^-_{j-\frac{1}{2}}\left|\Delta_{j-\frac{1}{2}}u\right| + \left(1 - C^-_{j+\frac{1}{2}} - C^+_{j+\frac{1}{2}}\right)\left|\Delta_{j+\frac{1}{2}}u\right| + C^+_{j+\frac{1}{2}}\left|\Delta_{j+\frac{1}{2}}u\right|\,. \tag{4.19b}$$

Summing over all j $(-\infty < j < \infty)$ gives the necessary conditions as the above equation must hold for all j. This takes the conservation principle into account resulting in the cancellation of most terms in the equations. □

Remark 16 *The theory of TVD schemes has also lead to implicit schemes based on these principles [110]. These have been used to produce steady-state profiles for aerodynamic designs in a variety of flow regimes [166]. In addition, the HOG and ENO [167] algorithms have also been extended to implicit time differencing. By taking the semi-discrete form of these equations*

$$\frac{\partial u}{\partial t} = C^+_{j+\frac{1}{2}}\Delta_{j+\frac{1}{2}}u - C^-_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u\,, \tag{4.20a}$$

with the conditions for a TVD approximation being

$$C^-_{j+\frac{1}{2}} \geq 0\,, \text{ and } C^+_{j+\frac{1}{2}} \geq 0\,. \tag{4.20b}$$

One can see that the set of linearly equations resulting from this scheme in the case of an implicit differencing is diagonally dominant and thus stable for solution by a variety of means.

Jameson and Lax [168] have provided a more general definition of a TVD scheme. This theorem provides conditions by which a scheme can have much larger support and be TVD. Shu [169] reports that Engquist and Osher had developed very high order TVD schemes along these lines.

Theorem 5 (Jameson and Lax [168]) *Given a semi discrete scheme*

$$\frac{du}{dt} = \sum_{k=-J\le j<J} C_j(k)\,\Delta_{j+\frac{1}{2}}u^n \,. \tag{4.21a}$$

is TVD if the following conditions are satisfied for all k

$$C_{-1}(k-1) \ge C_{-2}(k-2) \ge \ldots \ge C_{-J}(k-J) \ge 0 \tag{4.21b}$$

and

$$-C_0(k) \ge C_1(k+1) \ge \ldots \ge C_{J-1}(k+J-1) \ge 0 \tag{4.21c}$$

Remark 17 *This theorem when interpreted simply, means that the support for an interpolation within a given cell must decrease with distance from that point.*

Remark 18 *The questions relating to the stability and accuracy of a TVD approximation must be addressed separately from the question of its nature with regard to being TVD. It is often the case that when a scheme fails to provide TVD solutions, it also is essentially unstable.*

For instance, to prove a polynomial representation of a function is TVD (in one dimension), a general procedure can be defined. Taking the polynomial, $P_j(\theta)$ where

$$\theta = \frac{x - x_j}{\Delta x}$$

and then taking the case where $\lambda a > 0$. One defines

$$\tilde{P}_j(\theta) = P_j(\theta) - \ldots$$

with the function $\theta \in \left[-\frac{1}{2}, \frac{1}{2}\right]$. The form for the conservation law in proofs of TVD algorithms (explicit) is

$$u_j^{n+1} = u_j + C^+_{j+\frac{1}{2}}\Delta_{j+\frac{1}{2}}u^n - C^-_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u^n \,,$$

setting $C^+_{j+\frac{1}{2}} = 0$ then,

$$C^-_{j-\frac{1}{2}} = \lambda a\left[1 + Q_j(\theta) - Q_{j-1}(\theta)\right] \,,$$

where

$$Q_j(\theta) = \frac{\tilde{P}_j(\theta)}{\nabla_x u} \,.$$

The conditions to be TVD are

$$0 \le C^-_{j-\frac{1}{2}} \le 1 \,.$$

thus the following conditions can be brought to bear on the Q functions such that

$$Q_{j-1}(\theta) - Q_j(\theta) \leq 1 ,$$

and

$$Q_j(\theta) - Q_{j-1}(\theta) \leq \frac{1}{\nu} - 1 .$$

then the overall scheme is TVD while these are satisfied.

There are two major types of TVD schemes: the modified flux form [130] and the symmetric type [134]. The modified flux formulation is equivalent to a MUSCL type scheme for a [illegible] wave equation.

4.5.1 Modified Flux TVD Schemes

The modified flux TVD scheme has its dissipation function defined by

$$\phi^{MF}_{j+\frac{1}{2}} = \frac{1}{2}\left[g_j + g_{j+1} - \left|a_{j+\frac{1}{2}} + \gamma_{j+\frac{1}{2}}\right| \Delta_{j+\frac{1}{2}} u\right] , \tag{4.22a}$$

where

$$g_j = Q\left(\mu_{j-\frac{1}{2}} \Delta_{j-\frac{1}{2}} u . \mu_{j+\frac{1}{2}} \Delta_{j+\frac{1}{2}} u\right) , \tag{4.22b}$$

$$\gamma_{j+\frac{1}{2}} = \begin{cases} \dfrac{\Delta_{j+\frac{1}{2}} g}{\Delta_{j+\frac{1}{2}} u} & \text{if } \Delta_{j+\frac{1}{2}} u \neq 0 \\ 0 & \text{otherwise} \end{cases} . \tag{4.22c}$$

and

$$\mu = \frac{1}{2}\left(|a| - \lambda a^2\right) . \tag{4.22d}$$

4.5.2 Symmetric TVD Schemes

The symmetric TVD scheme has its dissipation function stated as

$$\phi^{SYM}_{j+\frac{1}{2}} = \left[\left(\left|a_{j+\frac{1}{2}}\right| - \lambda a^2_{j+\frac{1}{2}}\right) Q_{j+\frac{1}{2}} - \left|a_{j+\frac{1}{2}}\right|\right] \Delta_{j+\frac{1}{2}} u , \tag{4.23}$$

where $Q_{j+\frac{1}{2}}$ is a function of $\Delta_{j-\frac{1}{2}} u$, $\Delta_{j-\frac{1}{2}} u$, and $\Delta_{j+\frac{1}{2}} u$. The advantage of the symmetric TVD scheme is its lower cost in terms of arithmetic operations.

4.6 Flux-Corrected Transport

The flux-corrected transport scheme was the first algorithm developed that recognized the importance of Godunov's theorem. Some of the flux limiters (notably the minmod limiter) seem to have their genesis in the FCT method. Yet despite this, the other methods have flourished while the FCT methods have languished by comparison.

The original FCT was defined in a series of papers which gave analysis and results of using the scheme. The best recent reference is the book by Oran and Boris [4]. This method blends a high order flux with a low order monotone flux is such a way as to prevent the creation of new extrema. Although it is an improvement over classical methods, the FCT has not done well in tests against other modern algorithms [170, 44] and remains a pariah of sorts. The primary uses of the FCT have primarily been confined to turbulence [77], MHD [171] and reactive flow problems [172].

Zalesak [62] redefined the FCT in such a way as to make it more general. A standard low-order solution, similar to that obtained by donor-cell differencing, is used to define a monotonic solution. This solution is then used to limit an antidiffusive flux, which is defined as the difference between a high-order and low-order flux. As with the earlier versions of the FCT, the limiter is designed to give no antidiffusive flux when an extrema or a discontinuity is reached. This prescription of the FCT can allow the user to specify a wide range of low-order fluxes as well as a large variety of high-order fluxes. These have included central differencing of second- or higher-order, Lax-Wendroff, and spectral fluxes [173].

Recently, several researchers [174] have introduced an implicit FCT algorithm; however, this algorithm is limited to small multiples of the CFL number. This is because the low-order solution is produced by multiple sub-cycles with an explicit donor-cell (or other monotonic) solution and an implicit high-order solution. The high-order solution is only stable for small multiples of the CFL number, thus limiting the applicability of this algorithm. The FCT has also been extended for use with a finite-element solution method with great success [144][4].

One problem that plagues the FCT method is extension of the method to systems. Some schemes have used an equation-by-equation synchronization of flux limiters [144], but the results are not altogether pleasing. To my knowledge no one has published results of a Riemann solver being used to extend a FCT method to systems.

The flux-corrected transport algorithms can be written as follows:

1. find low-order monotonic cell-edge fluxes, $f^L_{j+\frac{1}{2}}$,
2. find the diffused solution, $\tilde{u}_j$,
3. find a high order flux $f^H_{j+\frac{1}{2}}$,
4. define an antidiffusive flux, $f^A_{j+\frac{1}{2}} = f^H_{j+\frac{1}{2}} - f^L_{j+\frac{1}{2}}$,
5. limit the antidiffusive flux to $f^C_{j+\frac{1}{2}}$, and
6. apply the corrected antidiffusive flux to the diffused solution to find u^{n+1}_j.

[4]The use of adaptive unstructured grids has been a key part of the success of this work.

The Boris and Book algorithm and Zalesak's algorithm differ only in a few steps. The Boris and Book algorithm uses a monotonic flux defined by

$$f^L_{j+\frac{1}{2}} = \frac{1}{2}(f_j + f_{j+1}) - \left(\frac{\lambda}{6} + \frac{\lambda}{3}a^2\right)(u_{j+1} - u_j) \ . \tag{4.24}$$

In Zalesak's algorithm, a simple donor-cell flux may be used (or any other monotone method) as the low-order flux. In the Boris and Book algorithm, the antidiffusive flux is defined by

$$f^A_{j+\frac{1}{2}} = \frac{1}{6}\left(\lambda - \lambda a^2\right)(u_{j+1} - u_j) \ , \tag{4.25}$$

and in Zalesak's algorithm it could be a Lax-Wendroff flux or another higher order flux minus the monotone flux.

Remark 19 *The formalism adopted above is from Zalesak's generalization. Boris and Book's original FCT was structured slightly differently, although the end result is equivalent. Their algorithm proceeds as follows [4, 175]: Compute a transported solution*

$$u^T_j = u^n_j - \lambda\left(f^T_{j+\frac{1}{2}} - f^T_{j-\frac{1}{2}}\right) \ . \tag{4.26a}$$

This solution is unstable and must be stabilized with a diffusion step

$$u^{TD}_j = u^T_j + \mu_{j+\frac{1}{2}}\left(u^T_{j+1} - u^T_j\right) - \mu_{j-\frac{1}{2}}\left(u^T_j - u^T_{j-1}\right) \ . \tag{4.26b}$$

This solution can then be corrected with an antidiffusion step, but this step is filtered with a flux limiter to avoid oscillatory solutions

$$u^{n+1}_j = u^{TD}_j + \nu^C_{j+\frac{1}{2}}\left(u^{TD}_{j+1} - u^{TD}_j\right) - \nu^C_{j-\frac{1}{2}}\left(u^{TD}_j - u^{TD}_{j-1}\right) \ , \tag{4.26c}$$

where ν is an antidiffusion coefficient not the CFL number.

Remark 20 *The main problem with the FCT is its lack of theoretical basis in the light of other modern methods. Were this present this method could move back toward the mainstream of numerical analysis.*

Before moving on, the results of the square wave test problem are given in Fig. 4.10. It should be noted that these results are very similar to those produced from the HOC algorithm (see Fig. 4.9). The results is somewhat less aesthetically pleasing due to a lack of symmetry. A similar test with a sine wave produces a "squaring" of the sine wave because of over compression

I explore FCT methods in a great deal of detail in Chapter 5, 6 and 7.

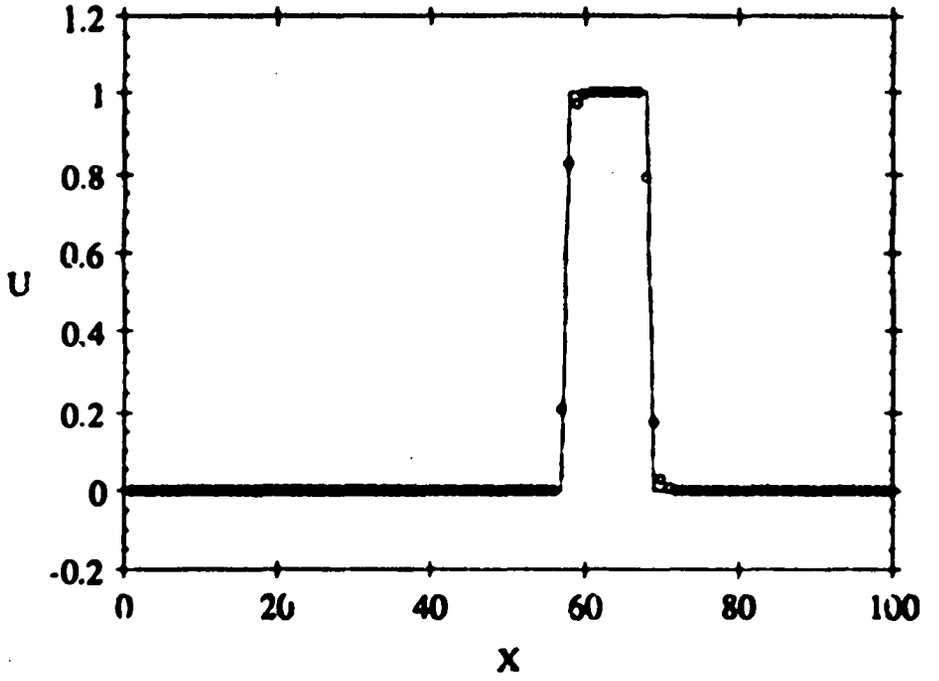


Figure 4.10: Computation of a square wave by the scalar wave equation using a FCT (Zalesak) algorithm.

4.7 The Role of Limiters

Flux, slope or gradient limiters play a pivotal role in the construction of modern methods for solving HCLs. The source of the nonlinearity necessary to produce high-order non-oscillatory algorithms is in these limiters. Despite their importance, the amount of work done toward understanding their behavior is relatively small [132, 176] and limited to a small class of schemes. A notable problem is that the analysis was confined to the same class of schemes, which are not necessarily representative of all the modern algorithms. This lapse in the collective understanding of limiters is important because limiters are a means through which a large class of modern numerical algorithms can be unified theoretically.

The FCT limiter has remained largely unstudied; the only major development is that of Zalesak [62]. The reasoning behind the form and function of the FCT limiter is unknown beyond the purely obvious. It is highly likely that both the FCT and other modern algorithms could benefit greatly from a greater understanding of their respective limiters.

At this point, it is useful to delineate the difference between slope and flux limiters more closely. This is done from the standpoint of a philosophical differentiation rather than from a purely technical basis. The slope limiters can be thought as being used directly during interpolation. Flux limiting usually involves methods that are classified as finite difference types. Thus slope limiting applies to HOC algorithms

and the flux limiting applies to TVD and FCT algorithms. One caveat can be placed on this classification; it is not stringent, an example of this are the ENO schemes due to Shu and Osher [65, 66].

A more complete description of limiters is given in Chapters 7 and 8.

4.8 The Role of Riemann Solvers

The role of Riemann solvers in modern methods for solving PDEs is not always clear. At one level, these method can be thought of as an essential ingredient for a successful algorithm, but at another level they appear to be a closure relation used to improve accuracy, or an extravagant feature which is not necessary.

The issue of Riemann solvers is critical to these types of methods. The philosophical basis of these methods is that the computational domain has been cut up into a number of discrete subdomains with the distinct possibility of discontinuities at the subdomain boundaries. The Riemann solvers resolve the behavior of the interaction of the subdomains. The Riemann solvers are integral parts of the schemes, but so is the fundamental differencing scheme. The prescription of the state of the fluid at the computational domain is as important (for high accuracy) as the solution for the ensuing fluid behavior. The Riemann solver however must ensure the physical nature (satisfaction of an entropy condition) of the solution.

Appendix B develops Riemann solvers in significantly more detail.

In the next chapter I begin the study of the design of high-resolution upwind shock-capturing methods through looking at the FCT method critcally.

Chapter 5.

An Improved Flux Corrected Transport Algorithm: A Finite Difference Formulation

---

> Iron rusts from disuse, stagnant water loses its purity, and in cold weather becomes frozen; even so does inaction sap the vigors of the mind. *Leonardo Da Vinci*

5.1 Introduction

As discussed before, Godunov [56] showed that the monotonic solution of first-order hyperbolic conservation laws is at most first-order accurate for linear differencing schemes. The first algorithm to successfully address this difficulty was the FCT algorithm of Boris, Book, and Hain [59, 140, 141, 142]. This algorithm performed quite well on linear advection problems and paved the way for future developments in the field. It essentially consisted of computing a solution with a nondiffusive transport method followed by a stabilizing diffusive step. This monotone solution is then used to aid in the construction of an antidiffusive step in which the solution from the first part of the algorithm is locally sampled and corrections are "patched" to it. This is accomplished with a flux limiter that only applies the flux corrections in the smooth part of the flow. As a result, the solution will be of a high-order in smooth parts of the convected profile, but first-order near discontinuities and steep gradients. Extension of the FCT algorithm to systems of conservation laws, however, has proved less successful.

Further developments on this topic were achieved by van Leer [60] in his higher order extensions of Godunov's method often referred to as MUSCL. The prescription of slope-limiting used by van Leer has great similarity to the flux-limiting used in the original FCT. The difficulties associated with FCT with systems equations is not shared by MUSCL because an exact solution to the local Riemann problem is used to construct the convective fluxes. While this approach adds complexity and cost to the solution procedure, the corresponding quality of the solution is greatly improved.

Zalesak [62] redefined the FCT in such a way as to make it more general. A standard low-order solution, similar to that obtained by donor-cell differencing, is used to define a monotonic solution. This solution is then used to limit an antidiffusive flux, which is defined as the difference between a high-order and low-order flux. As with the earlier versions of the FCT, the limiter is designed to give no antidiffusive flux when an

extrema or a discontinuity is reached. This prescription of the FCT can allow the user to specify a wide range of low-order fluxes as well as a large variety of high-order fluxes. These have included central differencing of second or higher order, Lax-Wendroff, and spectral fluxes [173]. Recently, several researchers [174] have introduced an implicit FCT algorithm; however, this algorithm is limited to small multiples of the CFL number. This is because the low-order solution is produced by multiple sub-cycles with an explicit donor-cell (or other monotonic) solution and an implicit high-order solution. The high-order solution is only stable for small multiples of the CFL number, thus limiting the applicability of this algorithm. The FCT has also been extended for use with a finite-element solution method with great success [144].

The performance of the explicit FCT algorithm is the subject of this chapter. Several investigators [170] [44] have noted for the older FCT algorithm that a lower CFL limit is required for stability. The FCT algorithm also suffers from being overcompressive (as is shown in Section 5.3). This was shown in a test of the FCT on a shock tube problem [143], where even at a CFL number of 0.1, the solution was of relatively poor quality. This probably is due to the handling of the pressure-related terms in the momentum and energy equations. This work aims to address these problems, first through making several improvements to the FCT and then by showing the extension of this modified FCT to systems of equations. In accomplishing this, I make extensive use of approximate Riemann solvers of the type introduced by Roe [63].

This chapter is organized into four sections. The following section provides an overview of the numerical solution of hyperbolic conservation laws. Later in that section, the FCT method according to Zalesak is introduced. This method is analyzed and suggestions for improvements are made including the extension of FCT to systems of equations. In the third section, results are presented for the methods discussed in this chapter. These results are for a scalar wave equation, Burgers' equation and a shock tube problem for the Euler equations. Finally, some closing remarks are made.

5.2 Method Development

The development of improved methods follows a short description of current FCT methods.

5.2.1 Zalesak's FCT Algorithm

Zalesak's FCT has been classified as a hybrid method that is applied in two steps. By being hybrid, the algorithm is based on the blending of high- and low-order difference schemes together. Step one is accomplished with a first-order monotonic solution such as donor-cell plus some additional diffusion (the entropy fix discussed in the previous section adds such dissipation). This could be accomplished with other first-order algorithms such as Godunov's [56] or Engquist and Osher's [127]. These fluxes are

used to produce a transported diffused solution $\tilde{u}$ as follows:

$$\tilde{u}_j = u_j^n - \sigma\left(f_{j+\frac{1}{2}}^{DC} - f_{j-\frac{1}{2}}^{DC}\right) . \tag{5.1}$$

A high-order flux, f^H, is defined in some way and then the low-order flux is subtracted from the high-order flux to define the antidiffusive flux as

$$f_{j+\frac{1}{2}}^{AD} = f_{j+\frac{1}{2}}^{H} - f_{j+\frac{1}{2}}^{L} .$$

The antidiffusive flux is then limited with respect to the local gradients of the conserved variable computed with the transported and diffused solution. Zalesak defined his limiter as a prelude to a truly multidimensional limiter, but also defined an equivalent limiter as

$$f_{j+\frac{1}{2}}^{C} = S_{j+\frac{1}{2}} \max\left\{0, \min\left[S_{j+\frac{1}{2}}\sigma^{-1}\Delta_{j-\frac{1}{2}}\tilde{u}, \left|f_{j+\frac{1}{2}}^{AD}\right|, S_{j+\frac{1}{2}}\sigma^{-1}\Delta_{j+\frac{3}{2}}\tilde{u}\right]\right\} , \tag{5.2}$$

where $S_{j+\frac{1}{2}} = \Delta_{j+\frac{1}{2}}\tilde{u}/\left|\Delta_{j+\frac{1}{2}}\tilde{u}\right|$ is the sign of the conserved variable's gradient spatially. This limiter is identical to the limiter defined by Boris and Book [59], but with a different definition of f^{AD}. The final cell-edge numerical diffusion is defined by

$$\phi_{j+\frac{1}{2}}^{FCT} = f_{j+\frac{1}{2}}^{C} + \phi_{j+\frac{1}{2}}^{DC} . \tag{5.3}$$

The FCT generally carries a stability limit on its time step of

$$\nu \leq 1 .$$

Before going further, several critical comments need to be made concerning this algorithm. Despite the striking generality, which is driven by the prescription of the antidiffusive fluxes, the algorithm has some deficiencies. By its formulation as a two-step method it has some disadvantages in terms of analytical analysis and efficiency of implementation. By the use of the inverse grid ratio σ^{-1} in the flux limiter, the algorithm is effectively limited to explicit time discretization (as is shown in the following section). The use of a diffused solution in the limiter is important in stabilizing the solution, which could yield oscillatory solutions without this step. Under closer examination, the use of a diffused solution acts as an upwind weighted artificial diffusion term. This sort of definition could lead to a fairly complex one-step FCT algorithm, which has, at first glance, similarity to UNO-type schemes. The diffusive terms in the FCT algorithm's limiter are upwind weighted rather than centered as with UNO based algorithms. Additionally, numerical experiments with a scalar advection equation show that the total variation for the FCT solution can increase with time for a CFL number less than one.

The use of higher order antidiffusive fluxes with this prescription of the FCT also

raises some questions about the actual order of the approximation. The antidiffusive flux is of the higher order, but the local gradients in the limiter are only accurate to second-order. This suggests that the solution may actually be of only second-order spatially (in the L_1 norm). This also holds for temporal order as the local gradient terms are only first-order in space, thus an antidiffusive flux based on a Lax-Wendroff flux may actually yield a first-order accurate temporal approximation. Thus the form of the local gradients used in the limiter may also need to be modified to accomplish the goal of true higher order accuracy.

5.2.2 A New FCT Algorithm

The first and simplest change is to rewrite the flux limiter as

$$f^C_{j+\frac{1}{2}} = S_{j+\frac{1}{2}} \max\left\{0, \min\left[S_{j+\frac{1}{2}}\tilde{\mu}_{j+\frac{1}{2}}\Delta_{j+\frac{1}{2}}\dot{u}, \left|f^{AD}_{j+\frac{1}{2}}\right|, S_{j+\frac{1}{2}}\tilde{\mu}_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}\dot{u}\right]\right\} , \qquad (5.4a)$$

where

$$\tilde{\mu}_{j+\frac{1}{2}} = \psi\left(\dot{a}_{j+\frac{1}{2}}\right) , \qquad (5.4b)$$

or

$$\tilde{\mu}_{j+\frac{1}{2}} = \psi\left(\dot{a}_{j+\frac{1}{2}}\right) - \sigma\dot{a}^2_{j+\frac{1}{2}} , \qquad (5.4c)$$

and $S_{j+\frac{1}{2}}$ has the same definition as before. See Section B.3.8 for the definition of ψ[1]. The second choice for $\tilde{\mu}_{j+\frac{1}{2}}$ gives second-order accuracy in both time and space if $f^{AD}_{j+\frac{1}{2}}$ is of similar or higher accuracy [61]. This relatively small change has a significant impact on the FCT algorithm, the solution is better behaved, and with some minor modifications can be stated as a stable implicit algorithm. This form is also a great deal closer to the definition of limiters used in TVD algorithms. However, this still leaves a two-step method which poses some problems from the standpoint of efficiency and extension to systems of conservation laws.

The similarities of this modification of the FCT with symmetric TVD schemes [134] are quite strong. The necessary changes to convert this scheme into one equivalent to the one described by Yee are simple. This consists of dividing the local gradient terms in the limiter by two and removing the first step of the FCT. Yee writes the numerical flux for the symmetric TVD method as

$$\dot{f}_{j+\frac{1}{2}} = \frac{1}{2}\left[a_{j+\frac{1}{2}}\left(u_j + u_{j+1}\right) - \psi\left(a_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u + Q_{j+\frac{1}{2}}\right] . \qquad (5.5)$$

An example of the $Q_{j+\frac{1}{2}}$ function would be

$$Q_{j+\frac{1}{2}} = S_{j+\frac{1}{2}} \max\left\{0, \min\left[S_{j+\frac{1}{2}}\psi\left(a_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u, \psi\left(a_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u,\right.\right.$$

[1] Essentially ψ is a smoothed definition of absolute value. The function is identical to the absolute value for most values, but is smoothed near the origin

$$S_{j+\frac{1}{2}}\psi\left(a_{j-\frac{1}{2}}\right)\Delta_{j-\frac{1}{2}}u\Big]\Big\}\ . \tag{5.6}$$

which strikes a strong resemblance with (5.4a) for an antidiffusive flux defined with a second-order central difference. For ease of analysis, this method is rewritten in the following form:

$$f_{j+\frac{1}{2}} = \frac{1}{2}\left[a_{j+\frac{1}{2}}\left(u_j + u_{j+1}\right) - \psi\left(a_{j+\frac{1}{2}}\right)\left(1 - Q_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u\right]\ , \tag{5.7}$$

where

$$Q_{j+\frac{1}{2}} = \text{minmod}\left(1, r^+_{j+\frac{1}{2}}, r^-_{j+\frac{1}{2}}\right)\ ,$$

with $r^+_{j+\frac{1}{2}} = \Delta_{j+\frac{1}{2}}u/\Delta_{j+\frac{1}{2}}u$ and $r^-_{j+\frac{1}{2}} = \Delta_{j-\frac{1}{2}}u/\Delta_{j+\frac{1}{2}}u$. The minmod limiter used with symmetric TVD schemes is defined by Yee, but has the same effect as (5.6). The minmod function of two arguments has the usual definition given in [45], which gives the same effect as the FCT limiter for three arguments. In words, the minmod limiter returns the minimum argument if the arguments are of the same sign and zero if the signs differ.

The FCT cell-edge flux can be written in the same way as the flux for a symmetric TVD scheme by defining

$$f^c_{j+\frac{1}{2}} = \frac{1}{2}\left|a_{j+\frac{1}{2}}\right|Q_{j+\frac{1}{2}}\Delta_{j+\frac{1}{2}}u\ , \tag{5.8}$$

if $Q_{j+\frac{1}{2}}$ is based on (5.4b)

$$Q_{j+\frac{1}{2}} = \text{minmod}\left(1, 2\tilde{r}^+, 2\tilde{r}^-\right)\ ,$$

and if $Q_{j+\frac{1}{2}}$ is based on (5.4c)

$$Q_{j+\frac{1}{2}} = \left(1 - \sigma\left|a_{j+\frac{1}{2}}\right|\right)\text{minmod}\left(1, 2\tilde{r}^+, 2\tilde{r}^-\right)\ ,$$

and

$$\tilde{r}^+ = \frac{\Delta_{j+\frac{1}{2}}\tilde{u}}{\Delta_{j+\frac{1}{2}}u}\ ,$$

$$\tilde{r}^- = \frac{\Delta_{j-\frac{1}{2}}\tilde{u}}{\Delta_{j+\frac{1}{2}}u}\ .$$

In [134] the inequalities that need to be satisfied in order for a flux of the form given in (5.5) to be TVD are

$$Q_{j+\frac{1}{2}} < 2\ . \tag{5.9a}$$

and

$$\frac{Q_{j+\frac{1}{2}}}{r^{\pm}_{j+\frac{1}{2}}} < \frac{2}{\sigma(1-\theta)\left|a_{j+\frac{1}{2}}\right|} - 2 , \tag{5.9b}$$

$$\nu < \frac{1}{1-\theta} , \tag{5.9c}$$

where θ is an implicitness parameter, such that $\theta = 0$ is fully explicit and $\theta = 1$ is fully implicit. The FCT limiter given in (5.4a) satisfies the first and last of these relations, but satisfaction of the other relation (5.9b) in a rigorous manner has proved to be more difficult. To establish some bounds on the properties of the FCT solutions, the first step of the FCT is ignored for the time being. Given this, the worst cases for the limiter are $Q = 2r^{\pm}$ or $2(1-\nu)r^{\pm}$. Comparing the first of these cases with (5.9b) gives

$$2 < \frac{2}{\sigma(1-\theta)|a|} - 2 ,$$

or

$$\nu < \frac{1}{2(1-\theta)} .$$

For the second of the two cases (only considered for $\theta = 0$),

$$2(1-\nu) < \frac{2}{\nu} - 2 ,$$

or

$$\nu < 1 .$$

Thus, even without the first step, the new FCT algorithm is TVD under some conditions. It is also unconditionally stable for fully implicit temporal discretization. The first step adds more dissipation into the algorithm, which should result in higher CFL limits for the first case. Numerical experiments confirm this and show that the new FCT is TVD for all CFL numbers less than one.

Zalesak's FCT can be subjected to a similar test after a reformulation of its limiter. Given the same definition as before for $f^{c}_{j+\frac{1}{2}}$,

$$Q_{j+\frac{1}{2}} = \left(1, \frac{2\tilde{r}^{+}}{\nu}, \frac{2\tilde{r}^{-}}{\nu}\right) , \tag{5.10}$$

where $\tilde{r}^{\pm}$ are defined as before. Using (5.9b), and again neglecting the first step, one can show that

$$\nu < \frac{\theta}{1-\theta} . \tag{5.11}$$

Thus, for a fully explicit approximation without the first step, Zalesak's FCT is never TVD. However, as the degree of implicitness increases, the algorithm becomes TVD for some CFL numbers and eventually becomes unconditionally TVD at $\theta = 1$. If one

looks at the form of the limiter as the CFL number increases, the effective antidiffusive flux reduces in an inversely proportional fashion. Therefore, at large CFL numbers, Zalesak's FCT is largely ineffective as a high-order implicit algorithm. Numerical experiments have shown that with the first step, Zalesak's FCT produces results that diminish in total variation up to a CFL number of about 0.95. Because the algorithm described above does not meet all my goals, further improvements are sought.

5.2.3 A Modified-Flux FCT Algorithm

To attain these goals, the FCT is recast in the form of Harten's modified- flux TVD scheme [61]. From this basis several FCT limiters can be shown to be TVD by the criteria given by [132], and the FCT can be written as a one-step method and extended to use as an implicit algorithm in the same way as TVD methods are [110]. This will be examined in the future.

The modified-flux TVD method is defined by computing cell-centered modified fluxes and making the overall flux upwind with respect to both the "physical" and modified fluxes. Formally, the modified-flux formulation has a dissipation term,

$$\phi^{MF}_{j+\frac{1}{2}} = \frac{1}{2}\left[g_j + g_{j+1} - \psi\left(a_{j+\frac{1}{2}} + \gamma_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u\right] , \tag{5.12a}$$

where

$$g_j = \text{minmod}\left(\mu_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u, \mu_{j+\frac{1}{2}}\Delta_{j+\frac{1}{2}}u\right) , \tag{5.12b}$$

and

$$\gamma_{j+\frac{1}{2}} = \begin{cases} \dfrac{\Delta_{j+\frac{1}{2}}g}{\Delta_{j+\frac{1}{2}}u} & \text{if } \Delta_{j+\frac{1}{2}}u \neq 0 \\ 0 & \text{otherwise} \end{cases} . \tag{5.12c}$$

A more general form of the minmod function is

$$\begin{aligned} \text{minmod}(a, b, n) = \text{sign}(a)\max\big[0, &\min(n|a|, \text{sign}(a)\,b), \\ &\min(|a|, n\,\text{sign}(a)\,b)\big] , \end{aligned} \tag{5.13}$$

which for $n = 2$ gives the Superbee limiter developed by Roe [176]. The function $\mu_{j+\frac{1}{2}}$ can have several forms, including

$$\mu_{j+\frac{1}{2}} = \frac{1}{2}\psi\left(a_{j+\frac{1}{2}}\right) . \tag{5.14a}$$

or

$$\mu_{j+\frac{1}{2}} = \frac{1}{2}\left[\psi\left(a_{j+\frac{1}{2}}\right) - \sigma a^2_{j+\frac{1}{2}}\right] . \tag{5.14b}$$

For (5.14a), the stability limit depends on the form of the limiter, for instance the

general minmod limiter yields a stability limit of

$$\nu \leq \frac{2}{(2+n)(1-\theta)}$$

for $n \leq 2$. The use of (5.14b) gives a stability limit of

$$\nu \leq 1$$

for all values of $n \leq 2$. The second definition has been recommended for explicit, time-accurate solutions [61] [110].

To formulate the FCT in a similar form, simply change the specification of the limiter. The traditional limiter used with the FCT is effectively a cell-edged flux rather than a cell-centered flux as needed for the modified-flux formulation. The definition of the antidiffusive flux must also be changed to a form more amenable to this formulation. This requires a more thoughtful statement of the antidiffusive flux, which can be easily incorporated with the type of formulation desired. For instance, the second-order central difference antidiffusive flux is

$$f^{AD}_{j+\frac{1}{2}} = \frac{1}{2}\psi\left(a_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u \,. \tag{5.15a}$$

or a Lax-Wendroff flux

$$f^{AD}_{j+\frac{1}{2}} = \frac{1}{2}\left[\psi\left(a_{j+\frac{1}{2}}\right) - \sigma a^2_{j+\frac{1}{2}}\right]\Delta_{j+\frac{1}{2}}u \,. \tag{5.15b}$$

or a fourth-order central difference

$$f^{AD}_{j+\frac{1}{2}} = \frac{1}{2}\psi\left(a_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u + \frac{1}{12}\left(\Delta_{j-\frac{1}{2}}f - \Delta_{j+\frac{3}{2}}f\right) \,. \tag{5.15c}$$

which can be written

$$f^{AD}_{j+\frac{1}{2}} = \frac{1}{2}\psi\left(a_{j+\frac{1}{2}}\right)\Delta_{j+\frac{1}{2}}u + \frac{1}{12}\left(a_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u - a_{j+\frac{3}{2}}\Delta_{j+\frac{3}{2}}u\right) \,.$$

These forms can be incorporated with a new limiter that has the desired properties. This limiter has the following form:

$$\begin{aligned}\text{minmod}(n) = S_{j+\frac{1}{2}}\max\Big|\,0, \min\Big(&\frac{1}{2}n\left|f^{AD}_{j+\frac{1}{2}}\right|, nS_{j+\frac{1}{2}}\mu_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u\Big), \\ &\min\Big(n\mu_{j+\frac{1}{2}}\left|\Delta_{j+\frac{1}{2}}u\right|, \frac{1}{2}nS_{j+\frac{1}{2}}f^{AD}_{j-\frac{1}{2}}\Big)\Big| \,,\end{aligned} \tag{5.16}$$

where $\mu_{j+\frac{1}{2}}$ is defined by (5.14a) or (5.14b).

Analysis of this limiter for the second-order central-difference-based antidiffusive flux follows that of Sweby [132]. For the values of $0 \leq n \leq 2$ in (5.16), the resulting

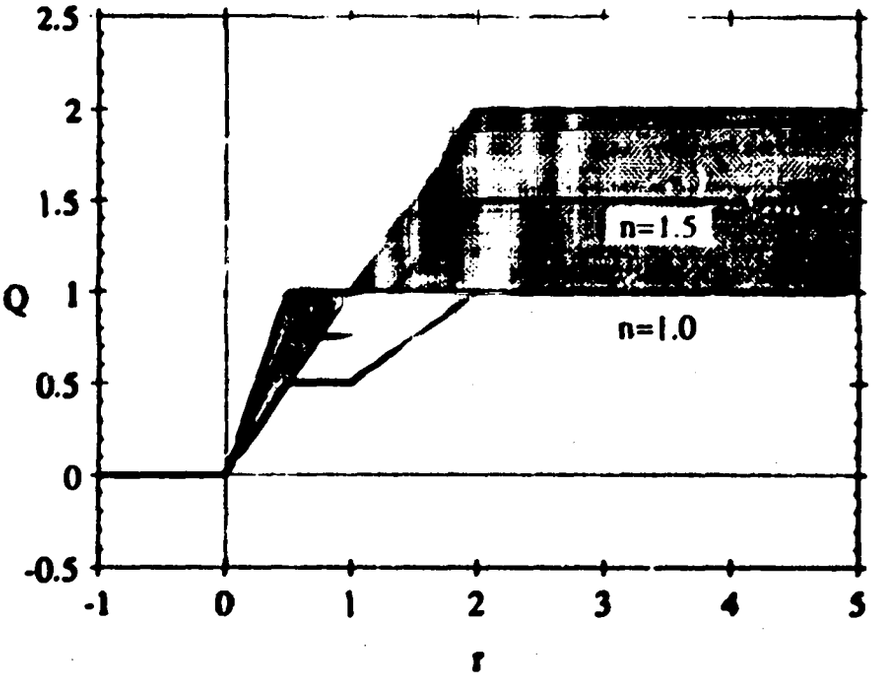


Figure 5.1: The characteristics of the FCT limiters for the modified-flux formulation.

limiter is in the TVD region of the curves shown in Fig. 5.1. For the value of $n = 2$, the resulting limiter is identical to Roe's Superbee limiter [176]. Shown in this figure are the plots for $n = 1$ and $n = 1.5$; the plot for $n = 2$ is identical to the upper boundary of the second-order TVD region. The boundaries of the second-order TVD region are shown by the thick lines on the plot. These limiters are second-order for all n for $r \leq 1/2$ and also second-order for $r \geq 2/n$. The only limiter of this class that is always second-order is the $n = 2$ limiter. The definition of r follows from Sweby's work.

5.2.4 Extension of FCT to Systems of Equations

The extension of the previously described methods to systems of hyperbolic conservation laws is no simple matter. The FCT currently is extended to systems in the simplest fashion. Traditional implementations of the FCT take the pressure terms in F as source terms and are handled with central differences. This leads to a poor representation of the wave interactions and the results that follow are often less than satisfactory.

The use of exact and approximate Riemann solvers offers a way through which more of the physical nature of the solution can be integrated into the solution procedure. To the authors' knowledge no attempt has been made to incorporate Riemann solvers with any of the previous FCT algorithms. Using van Leer's Riemann solver [60] [177], with Godunov's first-order method [56] [41] as the low-order method

with the first modification of the FCT, is my first attempt to incorporate a Riemann solver with FCT. While the results are better than the standard FCT implementation, they are worse than the Godunov method alone. To provide a more accurate and robust method, an approximate Riemann solver of the type introduced by Roe [63] is used.

The implementation of these Riemann solvers relies on the following transformations:

$$\Delta_{j+\frac{1}{2}} u^l = \sum_k r^k_{j+\frac{1}{2}} \alpha^k_{j+\frac{1}{2}} , \tag{5.17a}$$

where

$$\alpha^k_{j+\frac{1}{2}} = \sum_j l^k_{j+\frac{1}{2}} \Delta_{j+\frac{1}{2}} u^l . \tag{5.17b}$$

The numerical dissipation terms are then written as

$$\Phi^{DC}_{j+\frac{1}{2}} = \sum_k \frac{1}{2} r^k_{j+\frac{1}{2}} \psi\left(a^k_{j+\frac{1}{2}}\right) \alpha^k_{j+\frac{1}{2}} , \tag{5.18a}$$

$$\Phi^{FCT}_{j+\frac{1}{2}} = \sum_k r^k_{j+\frac{1}{2}} \left(f^{C\,k}_{j+\frac{1}{2}} + \Phi^{DC}_{j+\frac{1}{2}}\right) , \tag{5.18b}$$

and

$$\Phi^{MF}_{j+\frac{1}{2}} = \sum_k \frac{1}{2} r^k_{j+\frac{1}{2}} \left[g^k_j + g^k_{j+1} - \psi\left(a^k_{j+\frac{1}{2}} + \gamma^k_{j+\frac{1}{2}}\right) \alpha^k_{j+\frac{1}{2}}\right] , \tag{5.18c}$$

where

$$g^k_j = \text{minmod}\left(\mu^k_{j-\frac{1}{2}} \alpha^k_{j-\frac{1}{2}}, \mu^k_{j+\frac{1}{2}} \alpha^k_{j+\frac{1}{2}}\right) , \tag{5.18d}$$

and

$$\gamma^k_{j+\frac{1}{2}} = \begin{cases} \dfrac{\Delta_{j+\frac{1}{2}} g^k}{\alpha^k_{j+\frac{1}{2}}} & \text{if } \alpha_{j+\frac{1}{2}} \neq 0 \\ 0 & \text{otherwise} \end{cases} . \tag{5.18e}$$

Given these expressions for the numerical dissipation, the flux limiters used in the modified FCT (and for that matter classical FCT) Eqs. (5.2),(5.4a), and (5.16) are rewritten to take advantage of these forms. When a monotone first step is required with the FCT, Roe's first-order method [63] plus the entropy correction is used for the low-order method. The antidiffusive fluxes for the k^{th} wave are rewritten as

$$f^{AD}_{j+\frac{1}{2}} = \frac{1}{2} \psi\left(a^k_{j+\frac{1}{2}}\right) \alpha^k_{j+\frac{1}{2}} , \tag{5.19a}$$

or a Lax-Wendroff flux

$$f^{AD}_{j+\frac{1}{2}} = \frac{1}{2} \left[\psi\left(a^k_{j+\frac{1}{2}}\right) - \sigma \left(a^k_{j+\frac{1}{2}}\right)^2\right] \alpha^k_{j+\frac{1}{2}} . \tag{5.19b}$$

or a fourth-order central difference

$$f_{j+\frac{1}{2}}^{AD} = \frac{1}{2}\psi\left(a_{j+\frac{1}{2}}^{k}\right)a_{j+\frac{1}{2}}^{k} + \frac{1}{12}\left(a_{j-\frac{1}{2}}^{k}a_{j-\frac{1}{2}}^{k} - a_{j+\frac{1}{2}}^{k}a_{j+\frac{1}{2}}^{k}\right) . \qquad (5.19c)$$

For the classic FCT method, the flux limiter becomes

$$f_{j+\frac{1}{2}}^{C} = S_{j+\frac{1}{2}}\max\left[0, \min\left(\left|f_{j+\frac{1}{2}}^{AD}\right|, S_{j+\frac{1}{2}}\sigma^{-1}a_{j-\frac{1}{2}}^{k}, S_{j+\frac{1}{2}}\sigma^{-1}a_{j+\frac{1}{2}}^{k}\right)\right] . \qquad (5.20a)$$

The new FCT limiter becomes

$$f_{j+\frac{1}{2}}^{C\,k} = S_{j+\frac{1}{2}}\max\left[0, \min\left(\left|f_{j+\frac{1}{2}}^{AD}\right|, S_{j+\frac{1}{2}}\tilde{\mu}_{j-\frac{1}{2}}^{k}\tilde{a}_{j-\frac{1}{2}}^{k}, S_{j+\frac{1}{2}}\tilde{\mu}_{j+\frac{1}{2}}^{k}\tilde{a}_{j+\frac{1}{2}}^{k}\right)\right] ; , \qquad (5.20b)$$

where

$$\tilde{\mu}_{j+\frac{1}{2}}^{k} = \psi\left(\tilde{a}_{j+\frac{1}{2}}^{k}\right)$$

or

$$\tilde{\mu}_{j+\frac{1}{2}}^{k} = \psi\left(\tilde{a}_{j+\frac{1}{2}}^{k}\right) - \sigma\left(a_{j+\frac{1}{2}}^{k}\right)^{2} .$$

The modified-flux FCT method becomes

$$\begin{aligned}\text{minmod}\,(n) = S_{j+\frac{1}{2}}\max\Big[\,0, \min\Big(&\frac{1}{2}n\left|f_{j+\frac{1}{2}}^{AD}\right|, nS_{j+\frac{1}{2}}\mu_{j-\frac{1}{2}}a_{j-\frac{1}{2}}^{k}\Big),\\ &\min\Big(n\mu_{j+\frac{1}{2}}\left|a_{j+\frac{1}{2}}^{k}\right|, \frac{1}{2}nS_{j+\frac{1}{2}}f_{j-\frac{1}{2}}^{AD}\Big)\Big] ,\end{aligned} \qquad (5.20c)$$

where

$$\mu_{j+\frac{1}{2}}^{k} = \frac{1}{2}\psi\left(a_{j+\frac{1}{2}}^{k}\right)$$

or

$$\mu_{j+\frac{1}{2}}^{k} = \frac{1}{2}\left[\psi\left(a_{j+\frac{1}{2}}^{k}\right) - \sigma\left(a_{j+\frac{1}{2}}^{k}\right)^{2}\right] .$$

Again the FCT corresponding to the symmetric TVD schemes would require that (5.20b) be divided by two and the first step of the FCT removed from the algorithm. In the next section, the effects of these changes in the FCT is presented and compared with other standard methods.

It has come to my attention that Harten has developed similar ideas in [178]. These ideas are directly related to Harten's modified flux algorithm.

5.3 Results

To gauge the capability of the methods discussed in the previous sections, three test problems were solved with the FCT methods and several other high-resolution finite-difference methods. The other methods used are not described in detail here. The first test problem solves a scalar advection equation, on a uniform grid. Two problems are considered: a square wave and a sine wave over a complete period. Both waves

have an amplitude of one. The second problem is the inviscid Burgers' equation with initial data of a sine wave on a periodic domain with an amplitude of one. This solution is compared with the exact solution and the corresponding error norms are used to show convergence and order of approximation in these norms for the various methods. Finally, the shock tube problem used by Sod [41] is used as a vehicle for comparison of these methods for their use with systems of hyperbolic conservation laws.

The test problems are discussed in more detail in Appendix A. Specific differences in the use of the problems is given in the discussion.

5.3.1 Scalar Advection Equation

For the scalar advection of a square wave with a uniform velocity, the FCT performs quite well with very little numerical diffusion present in the solution. These solutions are obtained for a CFL number held constant at 1/2 after 80 time steps.

As shown in Fig. 5.2 (a), the square wave is captured quite well by the difference scheme, however, there is a distinct lack of symmetry in the solution. This lack of symmetry is evident in this version of the FCT despite the choice of the CFL number (which should lead to symmetric results, ideally). This can be attributed to the use of anti-upwind data by the limiter. This is more evident in Fig. 5.2 (b), but also evident is the overcompressive nature of the scheme. The sine wave is in the process of being compressed into two square waves. This behavior is clearly unacceptable because the character of the waves is largely destroyed by this algorithm. Figure 5.3 shows that the new FCT algorithm is somewhat more diffusive (less compressive) and has more of the expected symmetry in the solution. Figure 5.3 (b) still shows that this algorithm remains too compressive despite being TVD. One negative aspect of this calculation is the clipping of the extrema with respect to the previous figure, although overall this solution is superior in most respects to Zalesak's FCT.

By using the Lax-Wendroff fluxes as the base for the antidiffusive fluxes, the problem of overcompression is eliminated from both algorithms. This is at the cost of some clipping of the solution's extrema. The clipping in Fig. 5.4 is less than that in Fig. 5.5, but at the cost of the symmetry of the solution. The lack of symmetry is caused by the use of a computational velocity rather than a physical velocity in the limiter in Zalesak's FCT. Despite the dimensional consistency, this choice leads to incorrect local propagation speeds when the local gradients are chosen in the limiter, thus destroying the symmetry. The upwind bias is more evident in Zalesak's FCT, but is present in both solution techniques. This is caused by the first step of the FCT for Zalesak's algorithm, but in the new FCT, the use of the first step mitigates a lack of symmetry.

Figures 5.6 and 5.7 show the impact of the choice of n in the modified-flux FCT formulation (and for that matter other implementations of limiters). The lower value

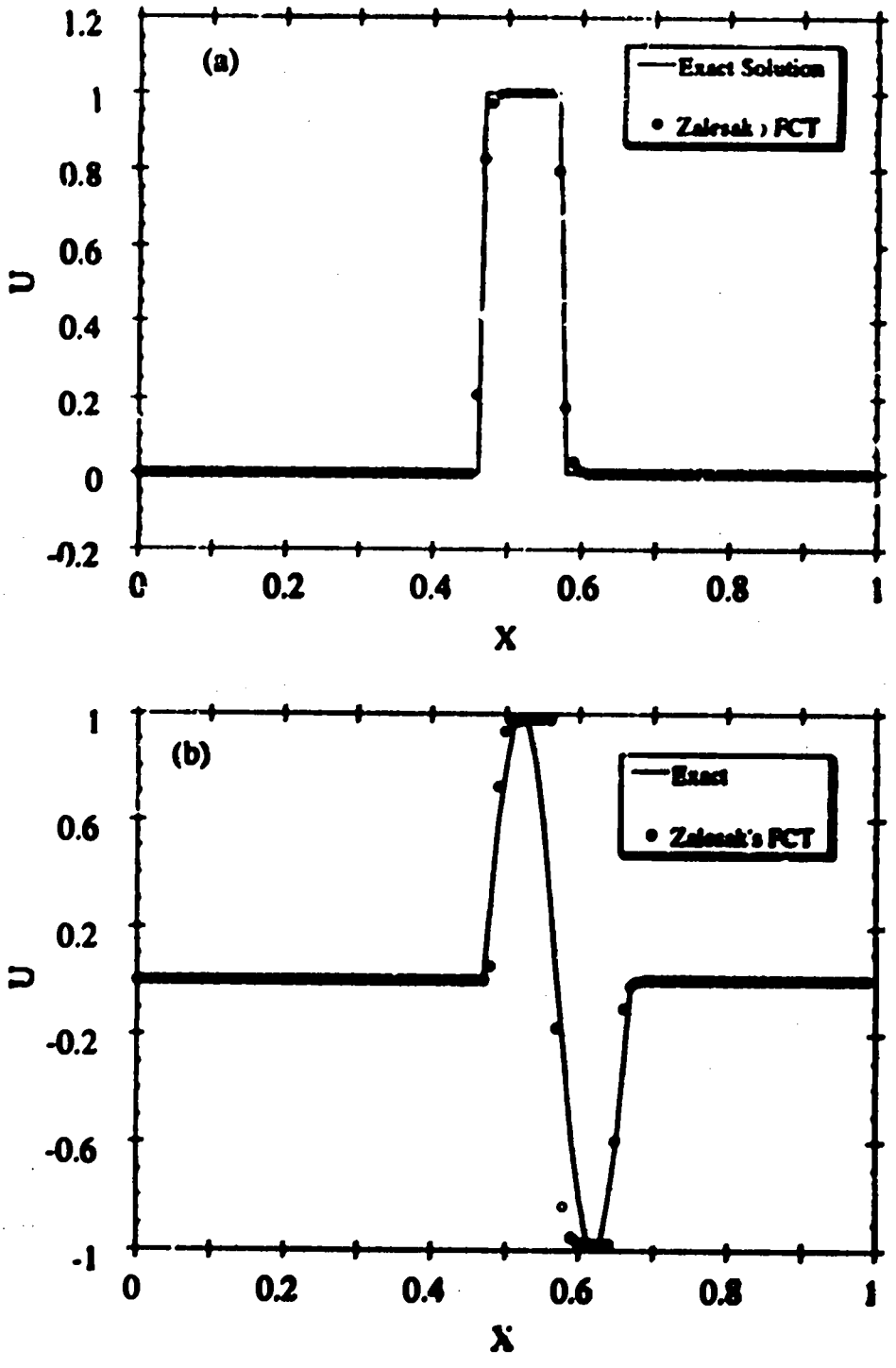


Figure 5.2: Solution of the scalar advection equation with Zalesak's FCT with the high-order flux defined by second-order central differencing.

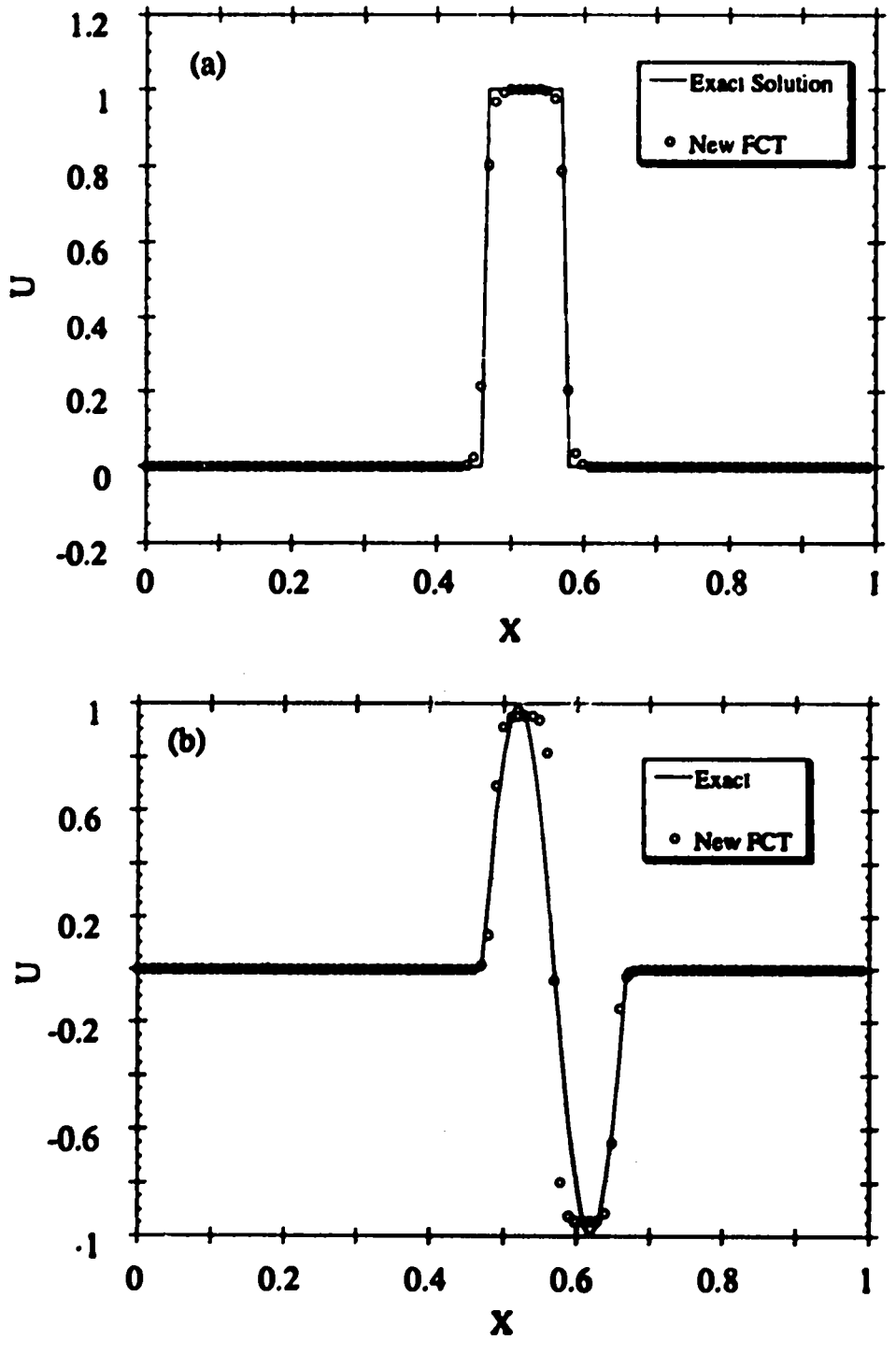


Figure 5.3: Solution of the scalar advection equation with the new FCT with the high-order flux defined Defined by second-order central differencing.

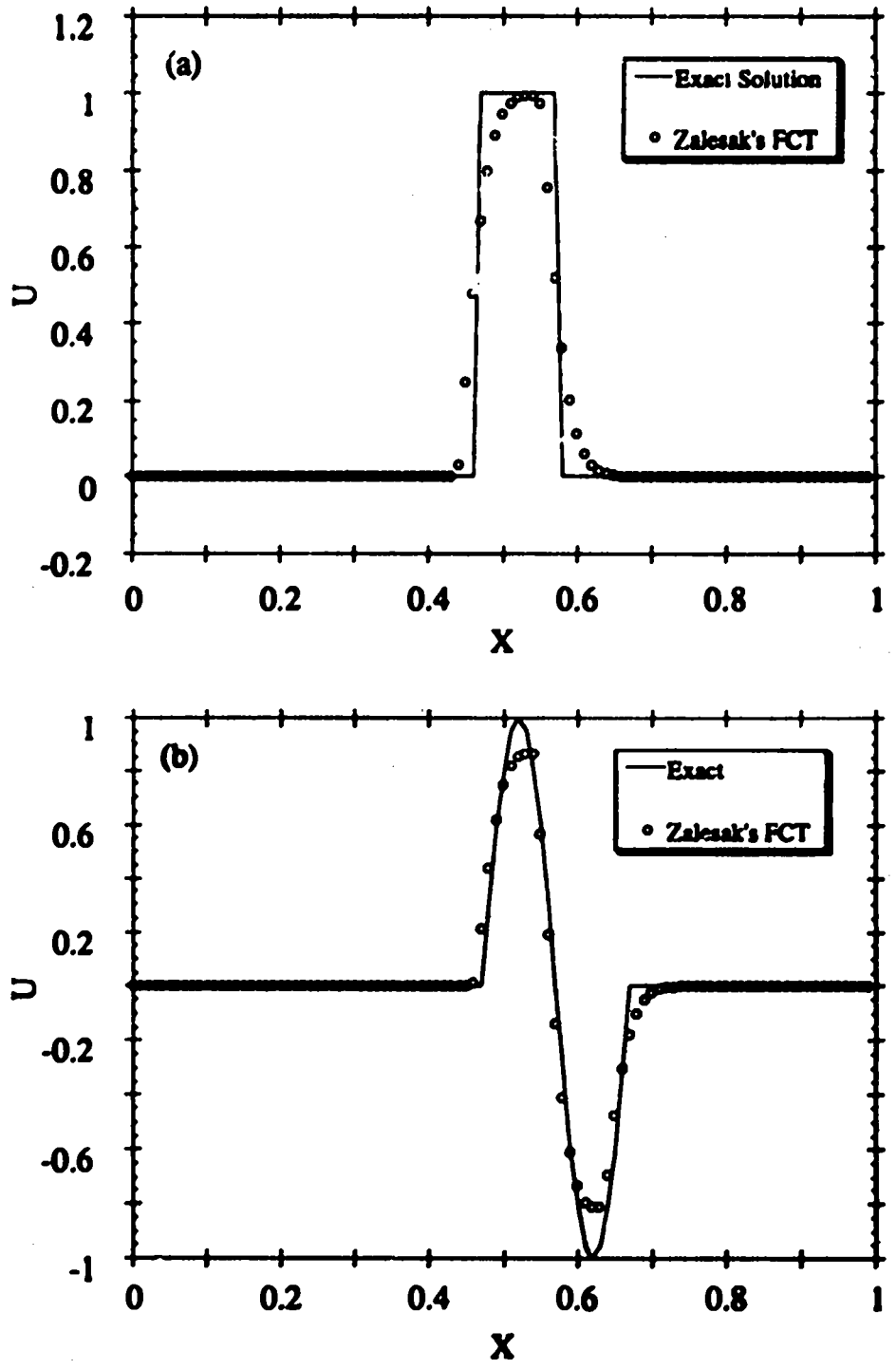


Figure 5.4: Solution of the scalar advection equation with Zalesak's FCT with the high-order flux defined defined by Lax-Wendroff differencing.

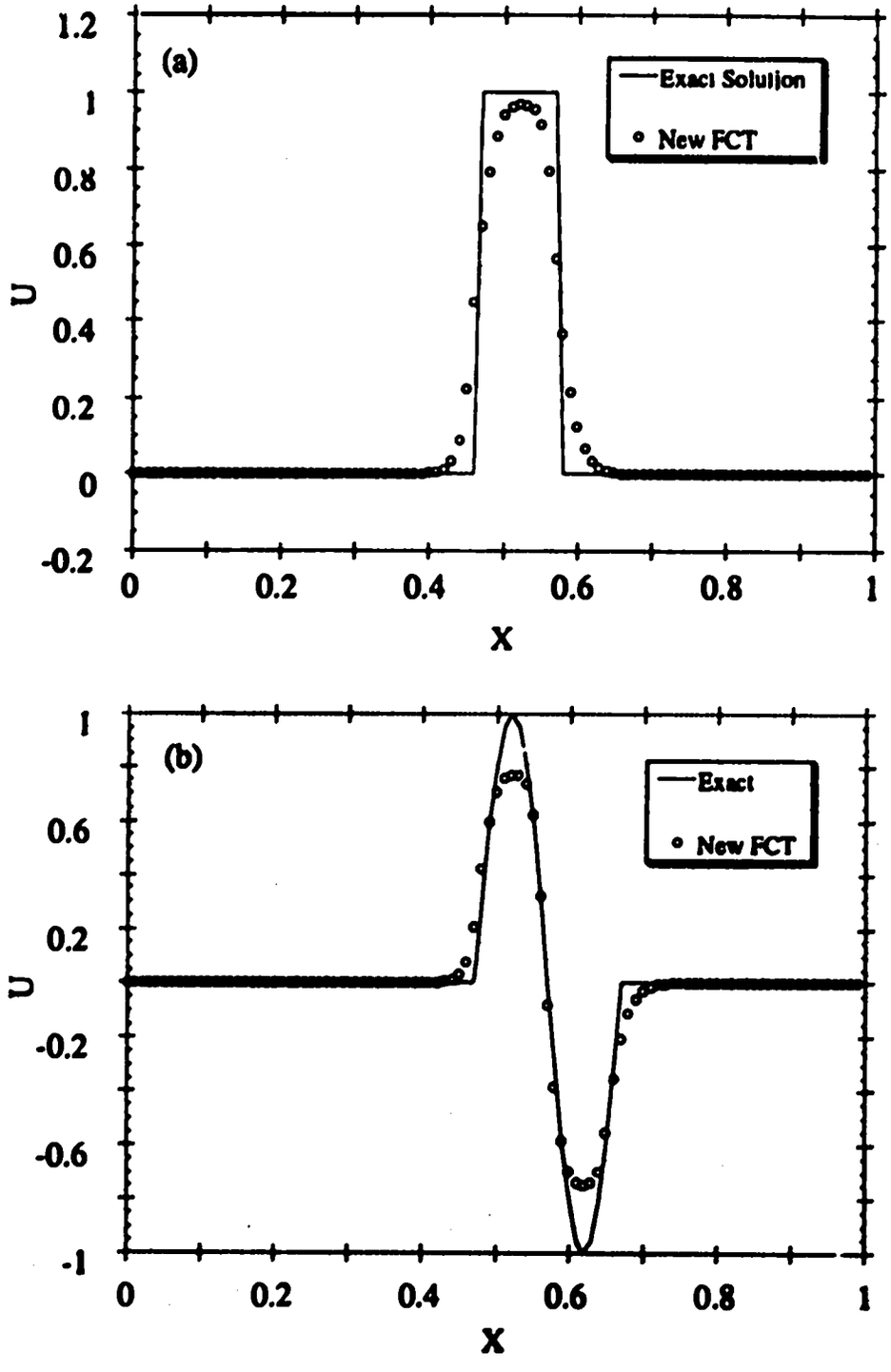


Figure 5.5: Solution of the scalar advection equation with the new FCT with the high-order flux defined by Lax-Wendroff differencing.

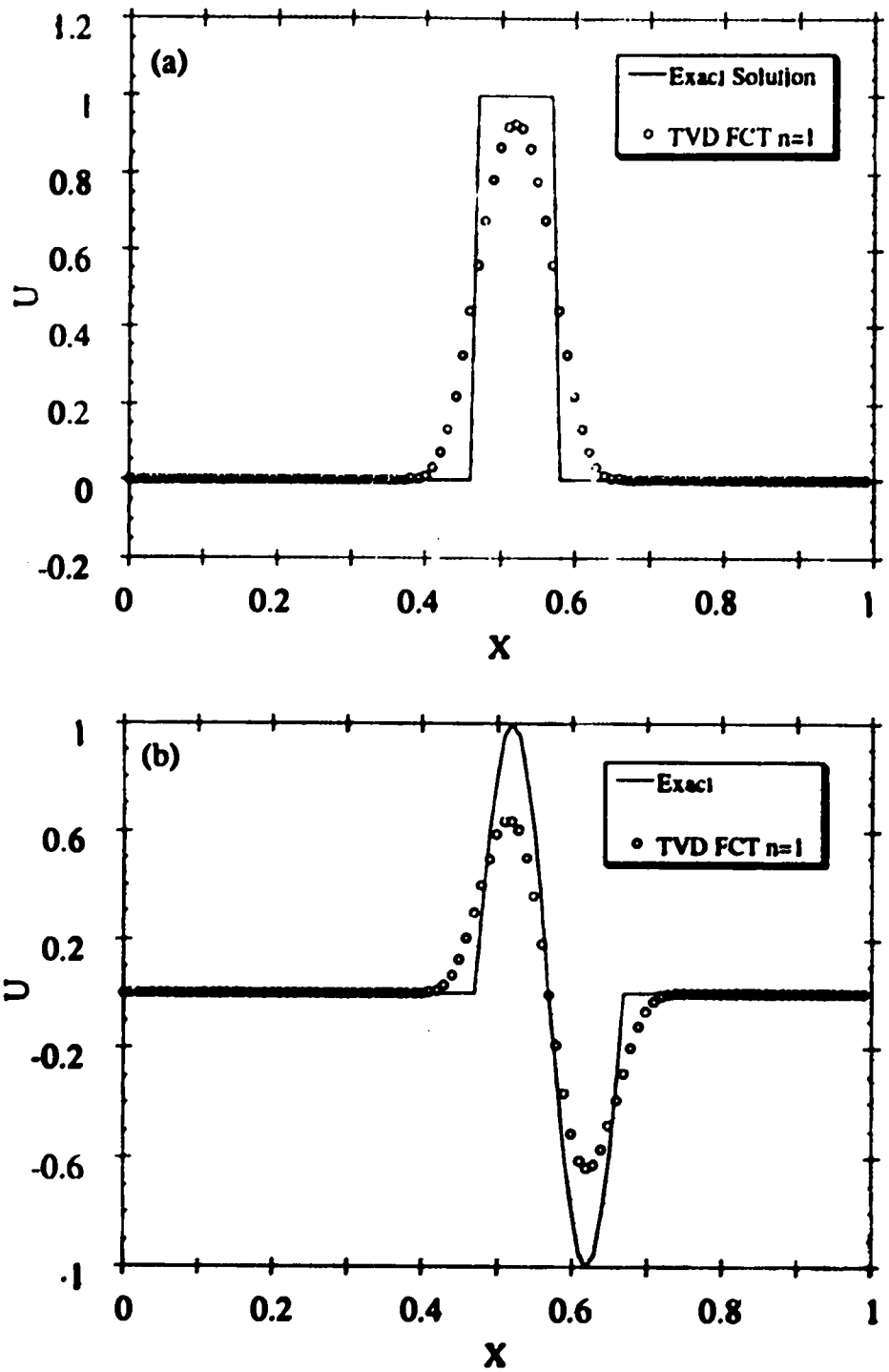


Figure 5.6: Solution of the scalar advection equation with the modified-flux FCT ($n = 1$ limiter).

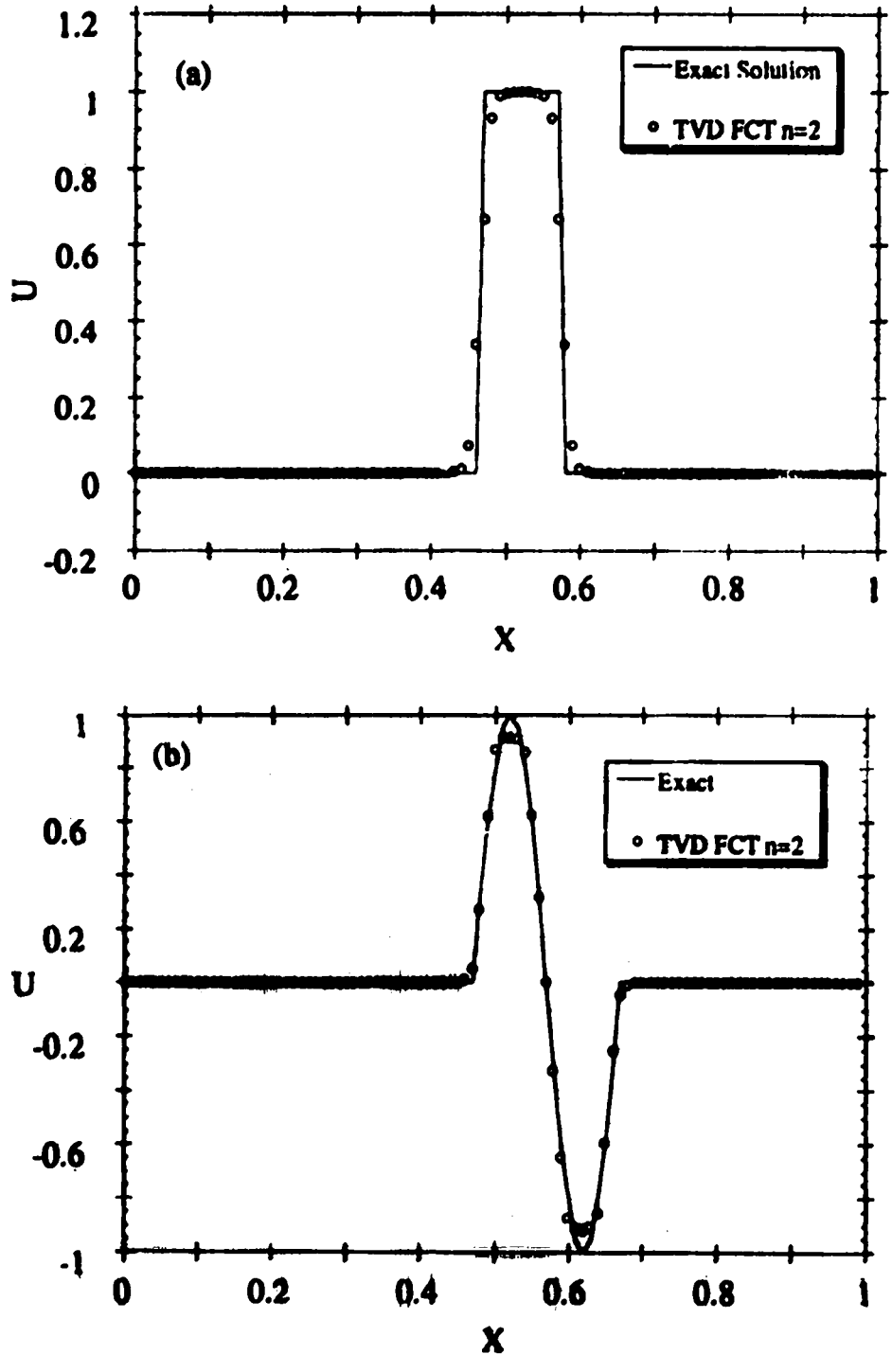


Figure 3.7: Solution of the scalar advection equation with the modified-flux FCT ($n = 2$ limiter).

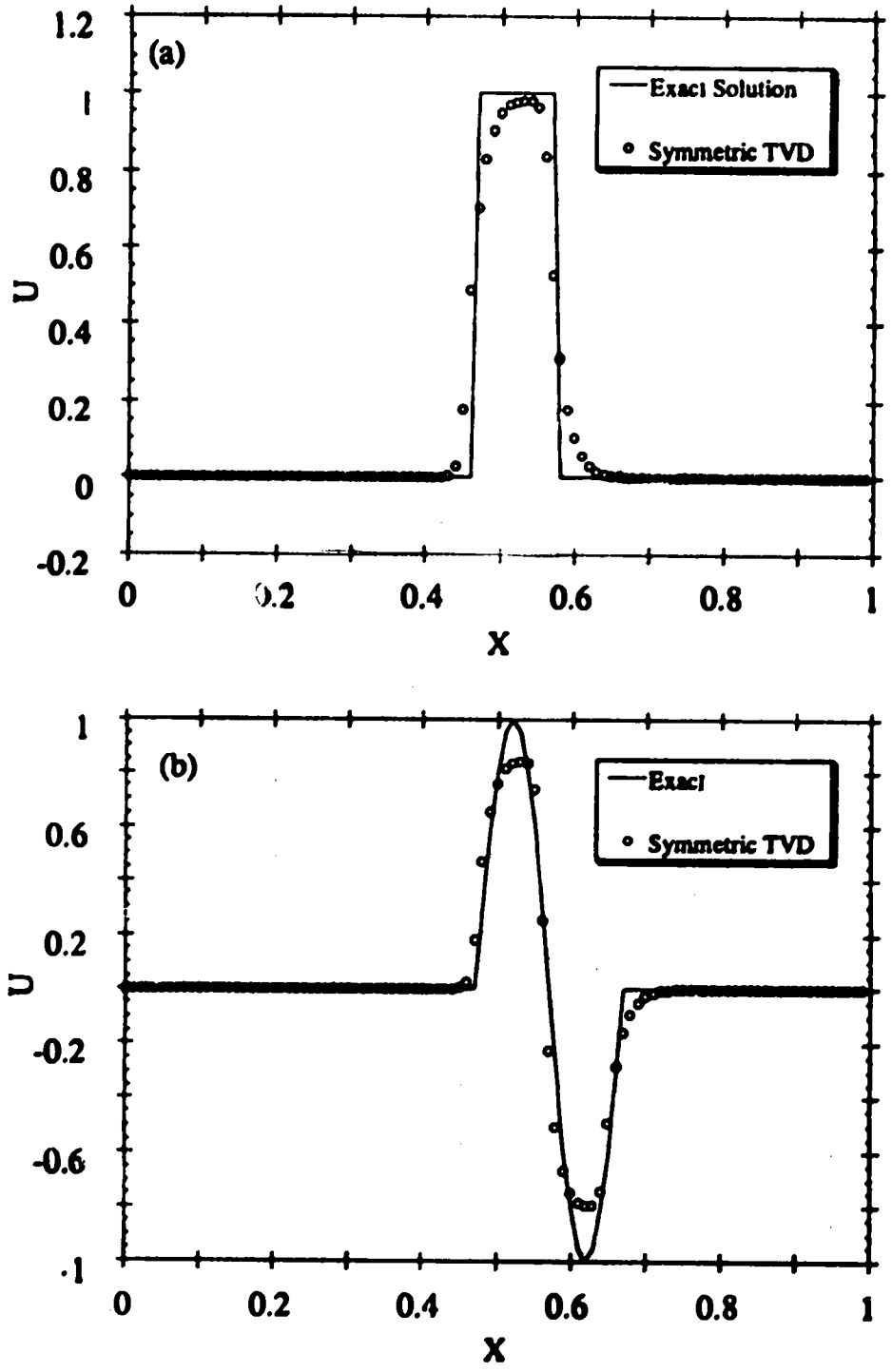


Figure 5.8: Solution of the scalar advection equation with a symmetric TVD scheme.

of n results in solutions that exhibit a great deal of dissipation and clipping of extrema. For the $n = 2$, solution is of high quality with the clipping of extrema quite controlled. This solution nearly equals that of the other FCT formulations for the square wave. For the sine wave, despite some clipping, the overcompression has disappeared with the character of the original profile well preserved.

The symmetric TVD algorithm (second-order in both time and space) produces results similar to the new FCT, but with a lack of symmetry. This can be cured with a predictive first step as with the FCT. As Fig. 5.8 shows, both exhibit a fair amount of extrema clipping and lack of symmetry. These are similar to the results obtained in Fig. 5.2 with Zalesak's FCT, but are more diffused.

5.3.2 Burgers' Equation

In all cases, the solutions obtained by using the high-resolution algorithms on Burgers' equation are quite good in terms of quality. Little would be gained by simply viewing their profiles (they are similar to the results in [110] for a TVD algorithm). By nature these high-resolution methods produce results that are first-order accurate in the L_∞ norm and approach second-order accuracy in the L_1 norm. In the next four figures discussed, figure (a) is for time equal to 0.2 when the solution remains smooth, and (b) shows the error norms (L_1, L_2 and L_∞) at time equal 1.0 after a shock has formed. For the methods used, each is second-order in time and space with the exception of the fourth-order FCT method, which is fourth-order in space. Second-order temporal accuracy is obtained by using a Lax-Wendroff type formulation. These calculations are all done with σ held constant.

In Fig. 5.9 the solution for $t = 0.2$ converges in the expected fashion, but at $t = 1$ problems are present with the convergence in the L_∞ norm. As the grid is refined, the L_∞ norm error increases rather than decreases as expected. As the grid size is further decreased convergence resumes, but is quite slow (about order 1/4). Figure 5.10 shows that the convergence properties of the fourth-order antidiffusive flux do not converge at a fourth-order rate and are in fact worse than those shown in the previous figure. The nonconvergence in the L_∞ norm for intermediate grid sizes for the $t = 1$ case is comparable. The new FCT algorithm shows slight improvements over both of these cases, but still has the same difficulties after a shock has formed in the solution. As shown by Fig. 5.11, the solutions converge faster than Zalesak's FCT, but are still plagued by some of the same problems. This behavior is also shared by the symmetric TVD's results in Fig. 5.3.2. The symmetric TVD does not converge as well as the new FCT method, but the nonconvergence problem is not as pronounced although it is clearly present.

The similarity of the solutions for the two FCT methods and the symmetric TVD algorithm, and the lack of such a problem in the modified-flux TVD method points to the form of the limiter as being the problem. The FCT and symmetric TVD use

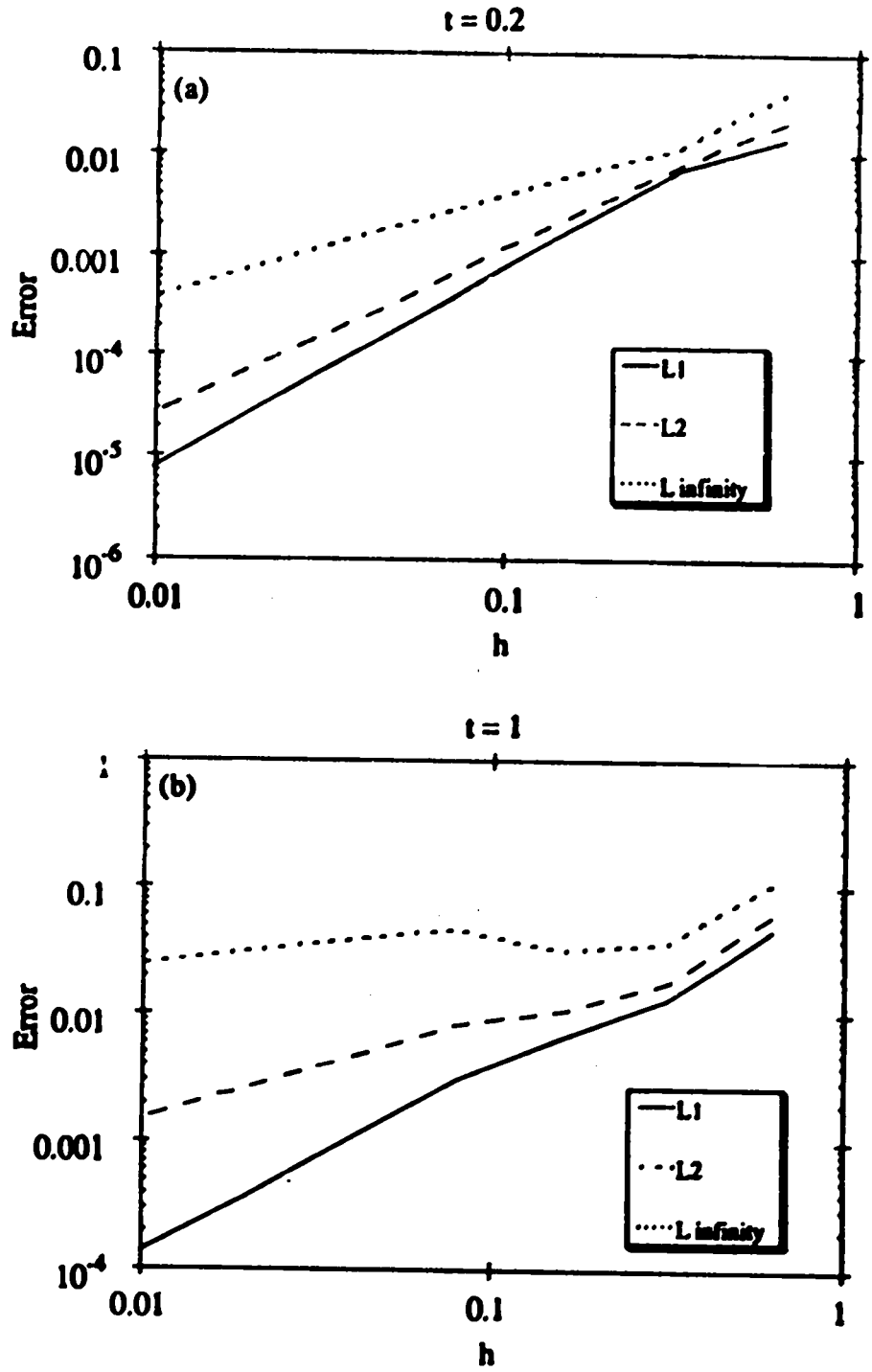


Figure 5.9: Convergence of error norms for Burgers' equation for Zalesak's FCT with the high-order flux defined by Lax-Wendroff differencing.

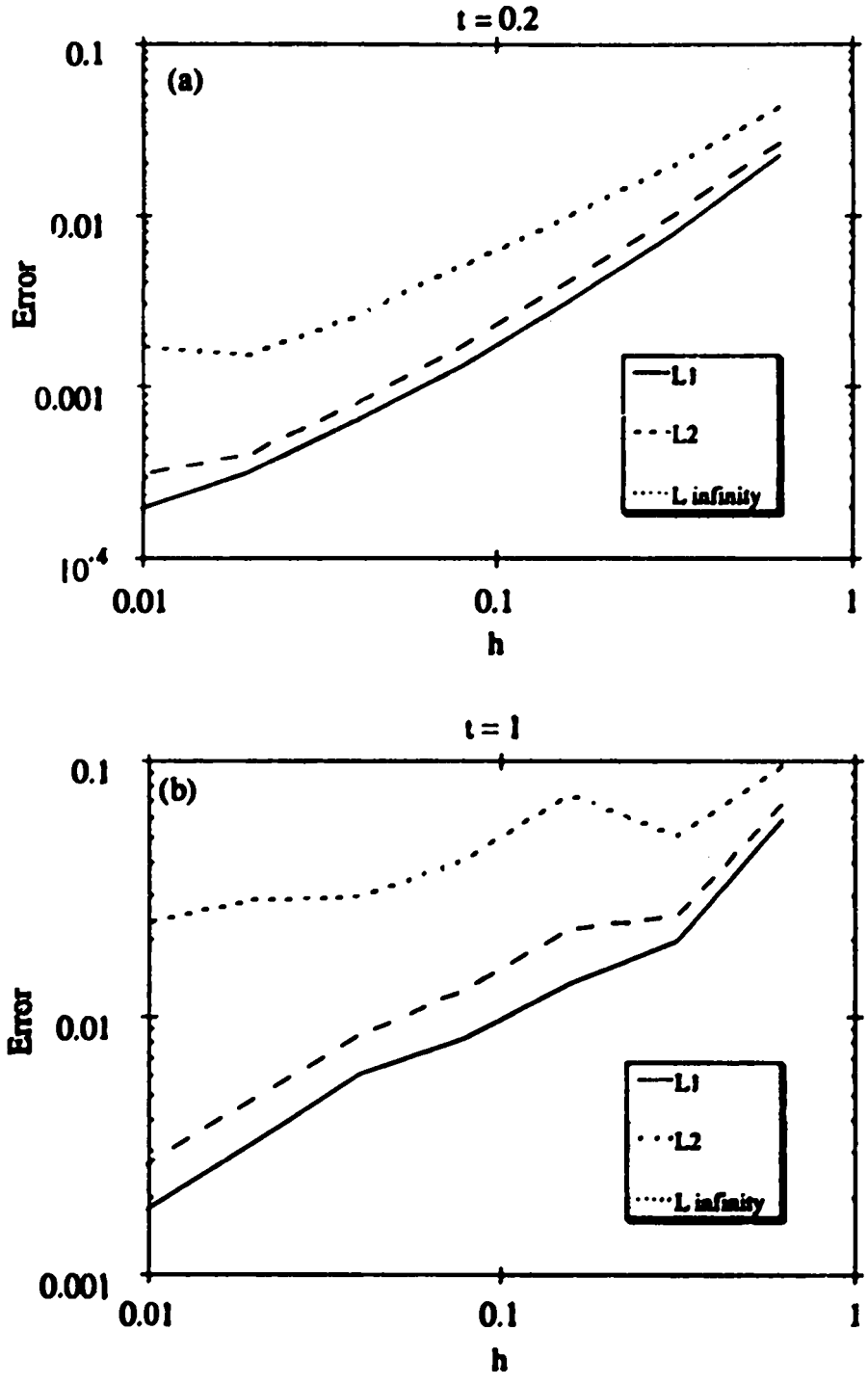


Figure 5.10: Convergence of error norms for Burgers' equation for Zalesak's FCT with the high-order flux defined by fourth-order central differencing.

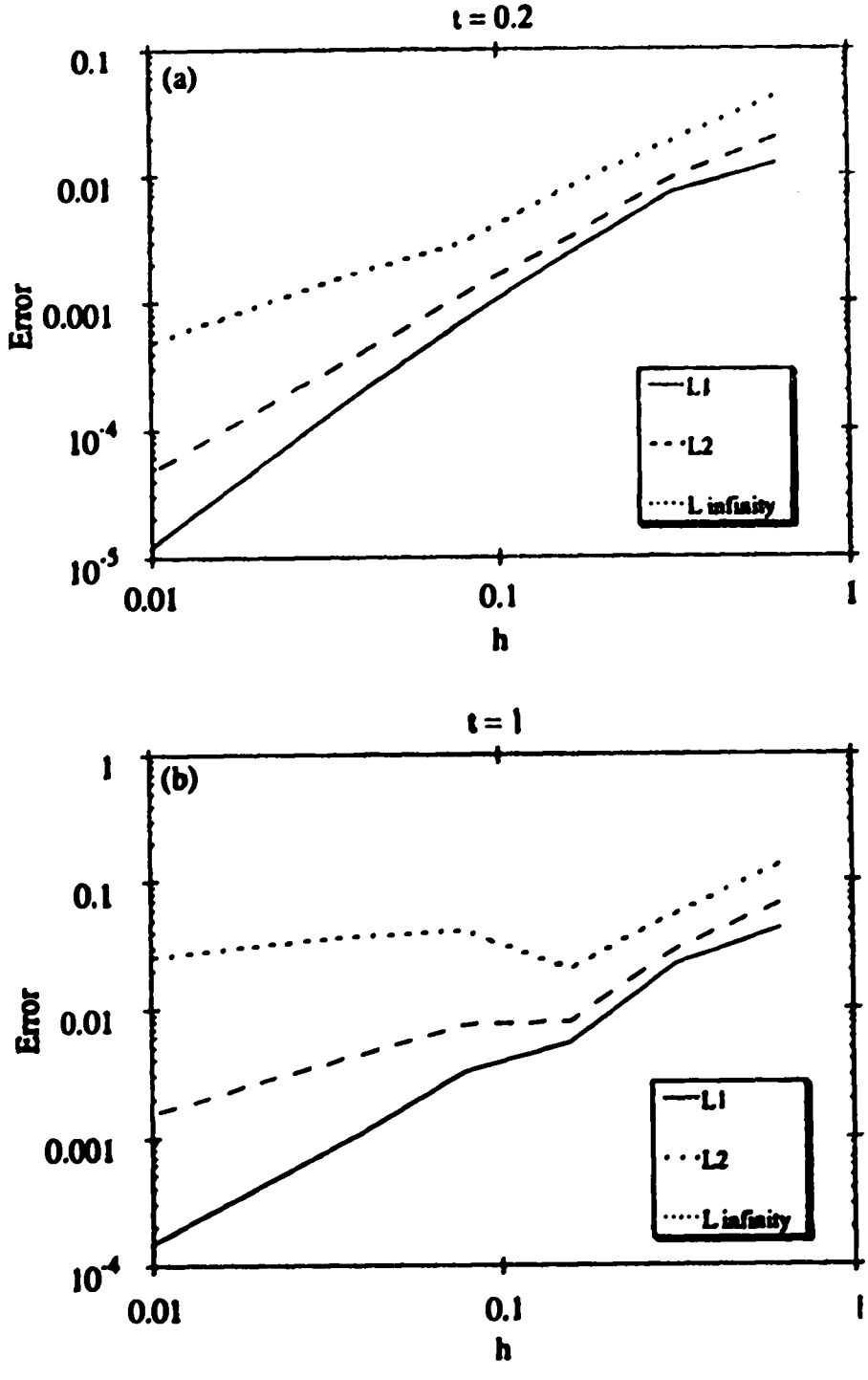


Figure 5.11: Convergence of error norms for Burgers' equation for the new FCT with the high-order flux defined by Lax-Wendroff differencing.

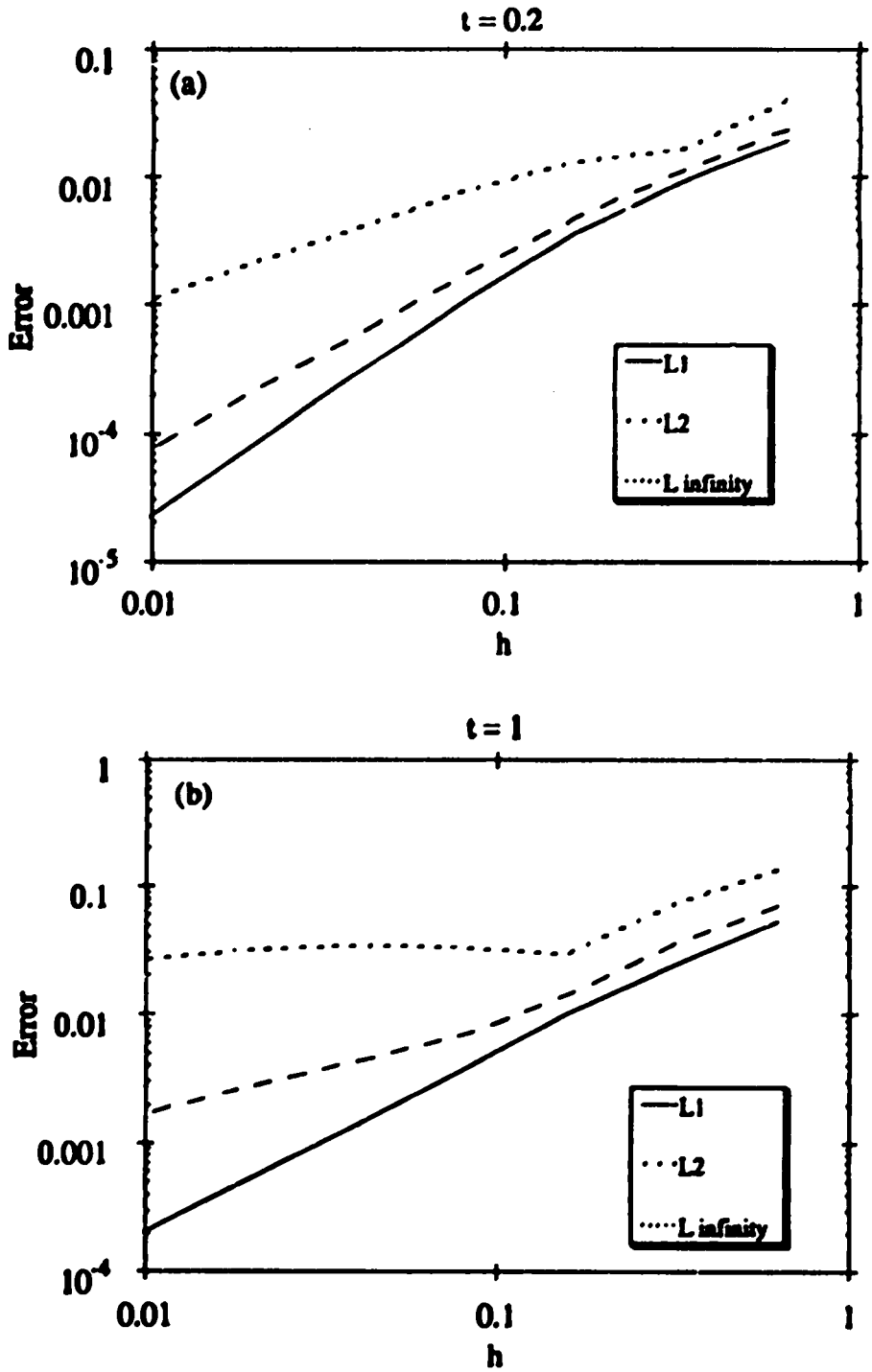


Figure 5.12: Convergence of error norms for Burgers' equation for a symmetric TVD algorithm.

cell-edged limiters rather than cell-centered limiters. This difference requires that each limiter has a wider spatial stencil than the cell-centered limiter, and as a result the resulting algorithm is not as sensitive to the presence of a discontinuity. This lack of sensitivity results in a poorer handling of shocks and discontinuities. The FCT is less diffusive than the symmetric TVD method, and this lack of diffusion increases the problem. The results for the fourth-order spatial limiter point out two problems: because the fourth-order spatial difference is more compressive than the second-order difference scheme, the convergence difficulty in the L_∞ norm at a shock is increased slightly. Experiments with a second-order Runge-Kutta time integration scheme show improvements in the L_1 convergence of the FCT.

5.3.3 Sod's Shock Tube Problem

The third problem involves the solution of Sod's test problem which tests the mettle of each algorithm against a difficult physical problem. For the FCT methods [in the modified-flux $\mu = 1/2(|a| - \sigma a^2)$], the Lax-Wendroff flux is used to define the antidiffusive flux. All results were produced for $\Delta t = 0.4\Delta x$ and shown for $t = 0.24$.

Figure 5.13 shows that the results using Zalesak's FCT are reasonable, but are polluted with a fair number of nonlinear instabilities. These instabilities are significantly worse if the limiter is based on a second- order central differences with numerous small expansion shocks present in the rarefaction fan. Even with the extra diffusion produced by the Lax- Wendroff flux, an expansion shock is present in the rarefaction wave and oscillations are present in the preshock region of the flow. The overall quality of this solution is quite poor. The new FCT formulation produces qualitatively better results that appear to be due to greater dissipation in the scheme. The expansion shock is no longer present. The overall quality of this solution is not high because of the considerable smearing of the features of the flow. In Fig. 5.14, the results show that a great deal of smearing is present except at the shock wave where the solution is very sharp. In both of these figures the pressure-related terms in the momentum and energy equations are incorporated as source terms rather than as convective fluxes, and are central differenced.

By computing the first step of the new FCT with Roe's first-order scheme, and using an approximate Riemann solver to compute the flux correction, the results are extremely good. As Fig. 5.15 shows, the smearing of a standard FCT implementation of the new FCT is gone, with the shock being computed with the same crispness. The rarefaction fan is smooth and in good agreement with the exact solution. The resolution of the contact discontinuity is somewhat smeared but is acceptable.

The modified-flux FCT (Fig. 5.16) has slightly poorer resolution of the contact discontinuity, but computes the shock in a sharper fashion. The overall quality of the solution is nearly identical to the previous case. In this case the value of $n = 1.5$ was used on all three fields. Better resolution of the contact discontinuity could be

obtained with the $n = 2$ limiter. The final two figures are shown for comparison with the previous figures. The symmetric TVD method (Fig. 5.17), gives adequate solution although the amount of smearing exceeds that of the other methods incorporating Roe's approximate Riemann solver. The UNO method (implemented with a method similar to the modified-flux TVD algorithm) was used to compute the solution shown in Fig. 5.18. This solution is of a quality similar to that found in Fig. 5.16 with slightly better resolution of each of the features of the flow.

5.4 Concluding Remarks

The modifications proposed in this work on the FCT algorithm of Zalesak have proved to be quite successful in terms of performance and in terms of yielding a better understanding of the FCT algorithm in general. These modifications give an algorithm that is formally second-order in both time and space. Also, the extension of this method to systems of equations is a good deal more effective than the typical extension of the FCT to systems. The notion that the FCT algorithm for certain cases may be TVD (subject to certain restrictions on the CFL number) is quite gratifying. It is perhaps more useful to consider the flexibility of the formulation of this FCT with respect to a wider range of high-order fluxes. This gives the prospect of formulating solutions that have higher orders of approximation than previously attempted and also have a reasonable extension to systems of equations.

Future work includes the modification of the FCT to include MUSCL-type schemes as well as the appropriate generalization of Zalesak's multidimensional limiter to these types of methods. As mentioned earlier, these methods, once cast in the appropriate form, can be used for implicit time integration where the necessary form is similar to that found in TVD implicit formulations. Tests on simple test problems indicate that these methods are unconditionally stable.

The initial motivation of this work was to tie together in a more coherent fashion the various modern high-resolution methods for numerically solving hyperbolic conservation laws. This work should be considered a start, with the advances mentioned above, as progress toward this goal.

The next chapter explores the topic of this chapter further. The link between flux-corrected transport and high-order Godunov schemes is shown and explored further.

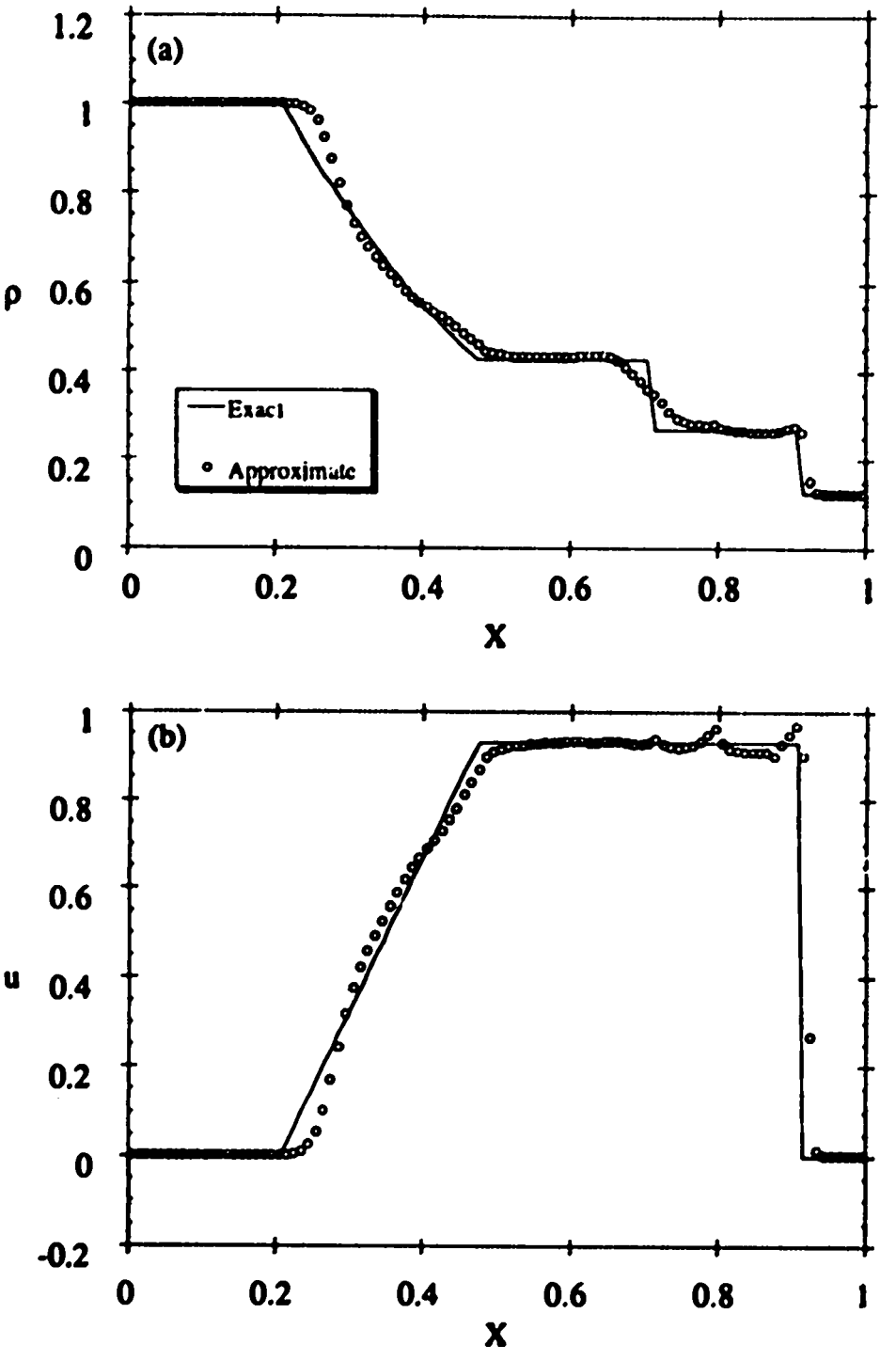


Figure 5.13: Solution of Sod's shock tube problem with Zalesak's FCT.

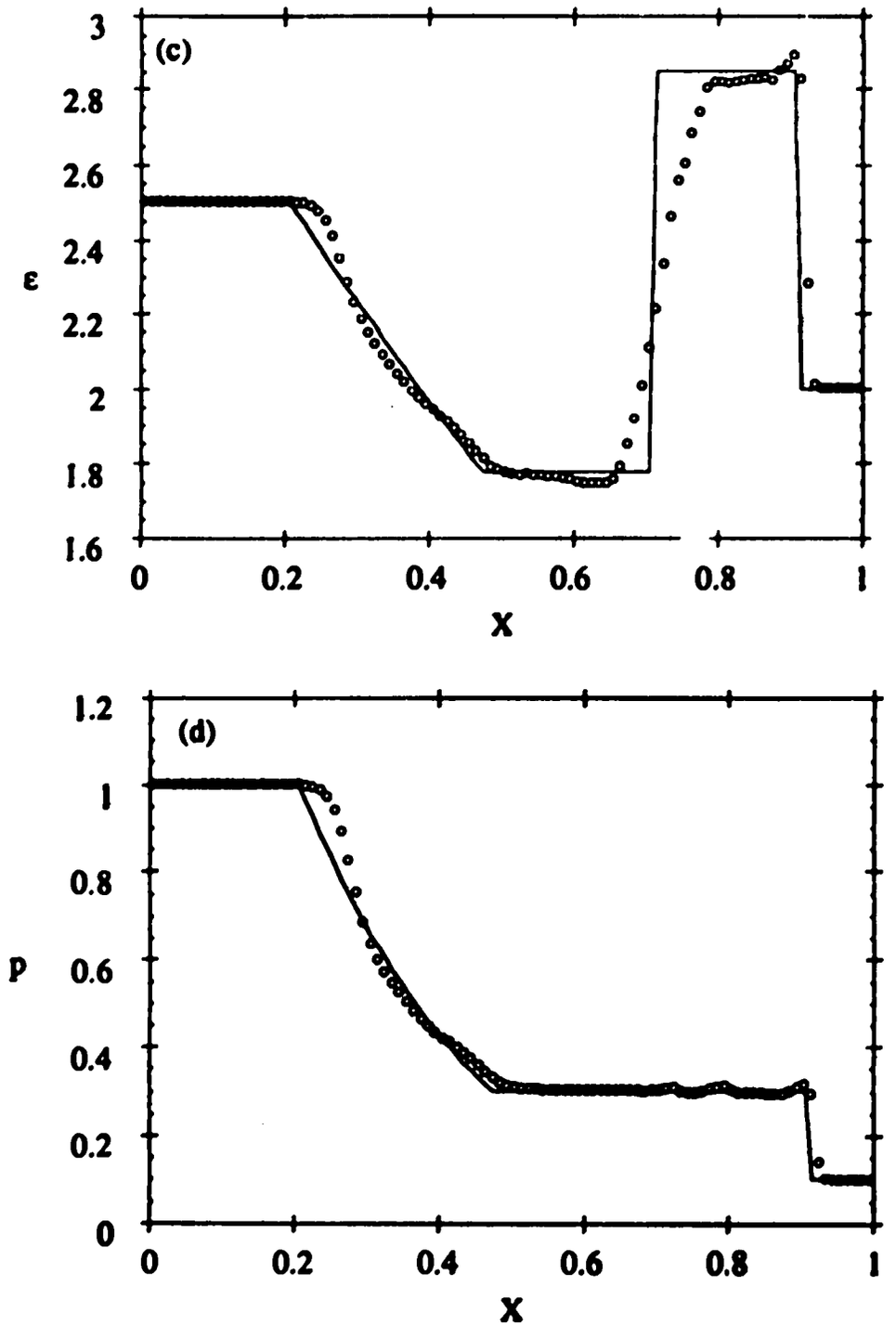


Figure 5.13: continued.

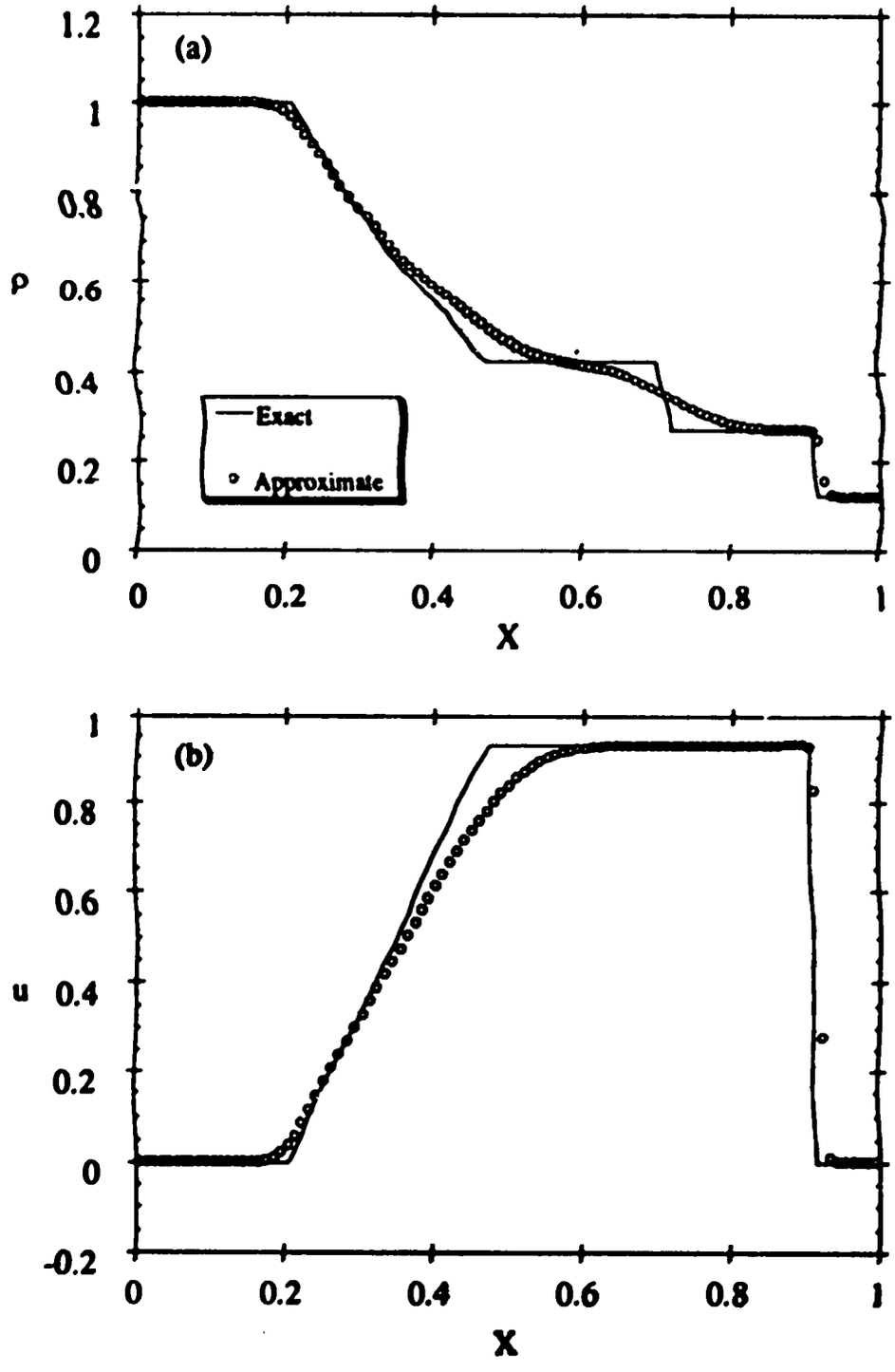


Figure 5.14: Solution of Sod's shock tube problem with the new FCT.

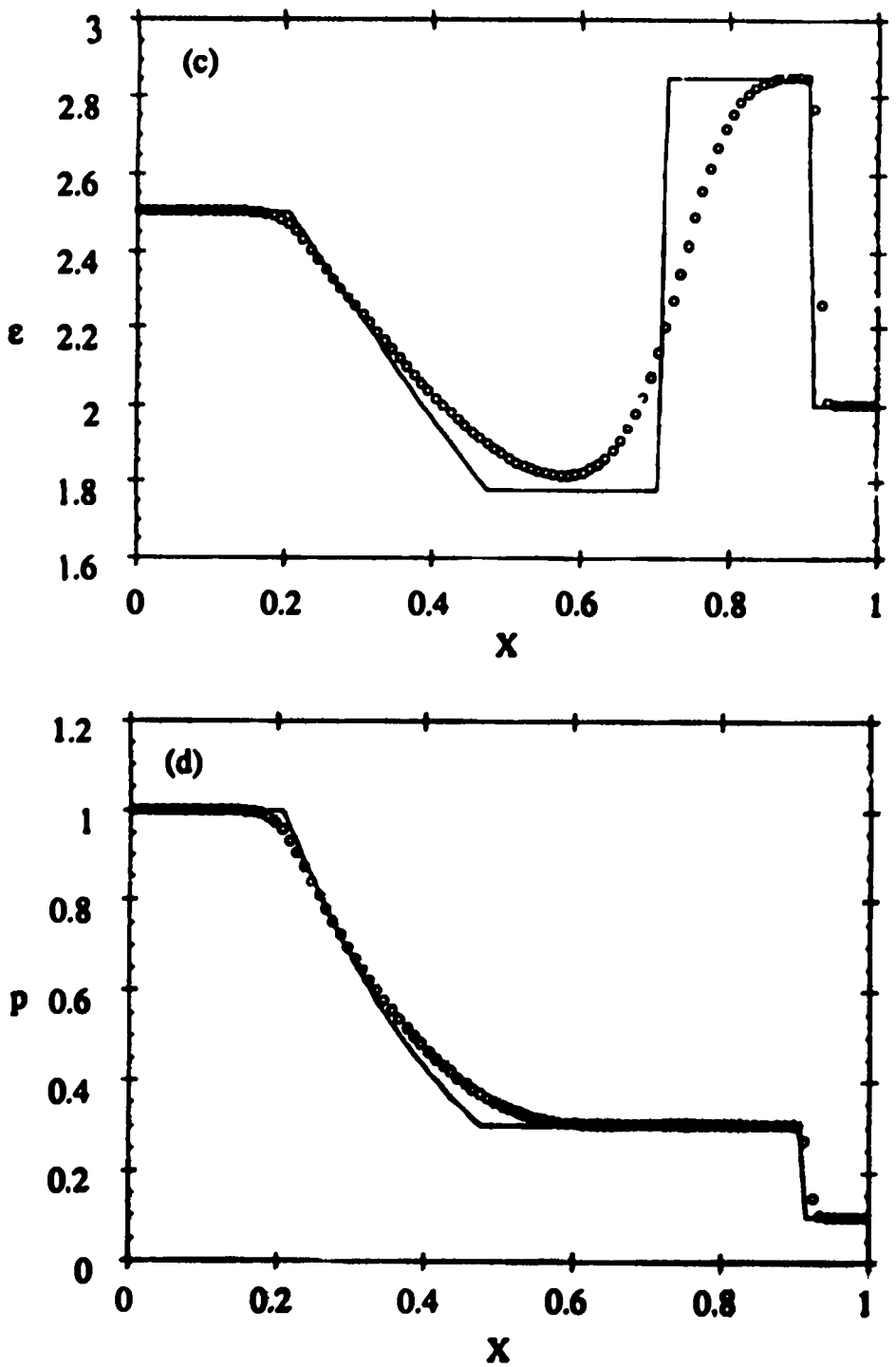


Figure 5.14: continued.

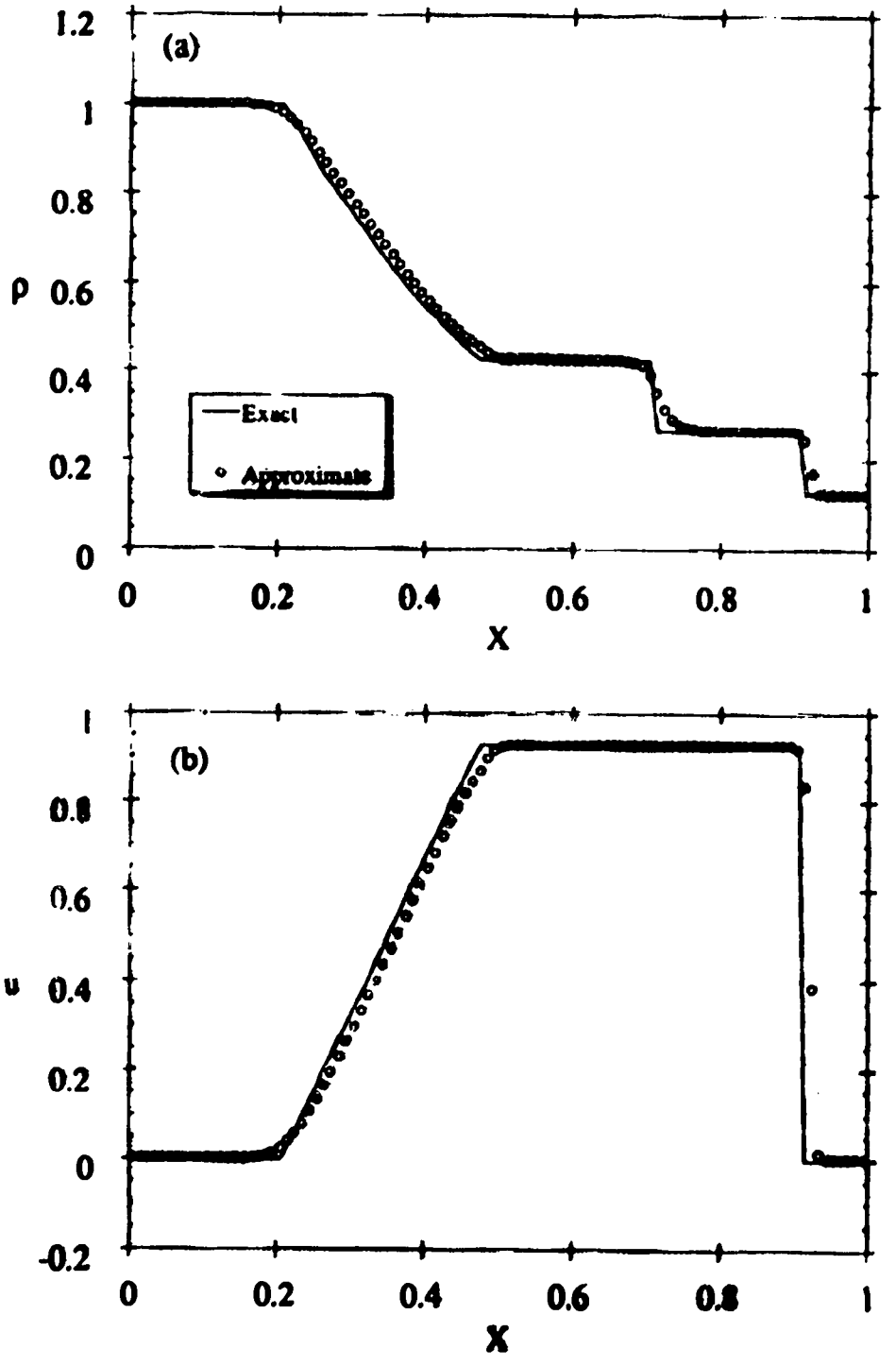


Figure 5.15: Solution of Sod's shock tube problem with new FCT with Roe's approximate Riemann solver used to define both low- and high-order fluxes.

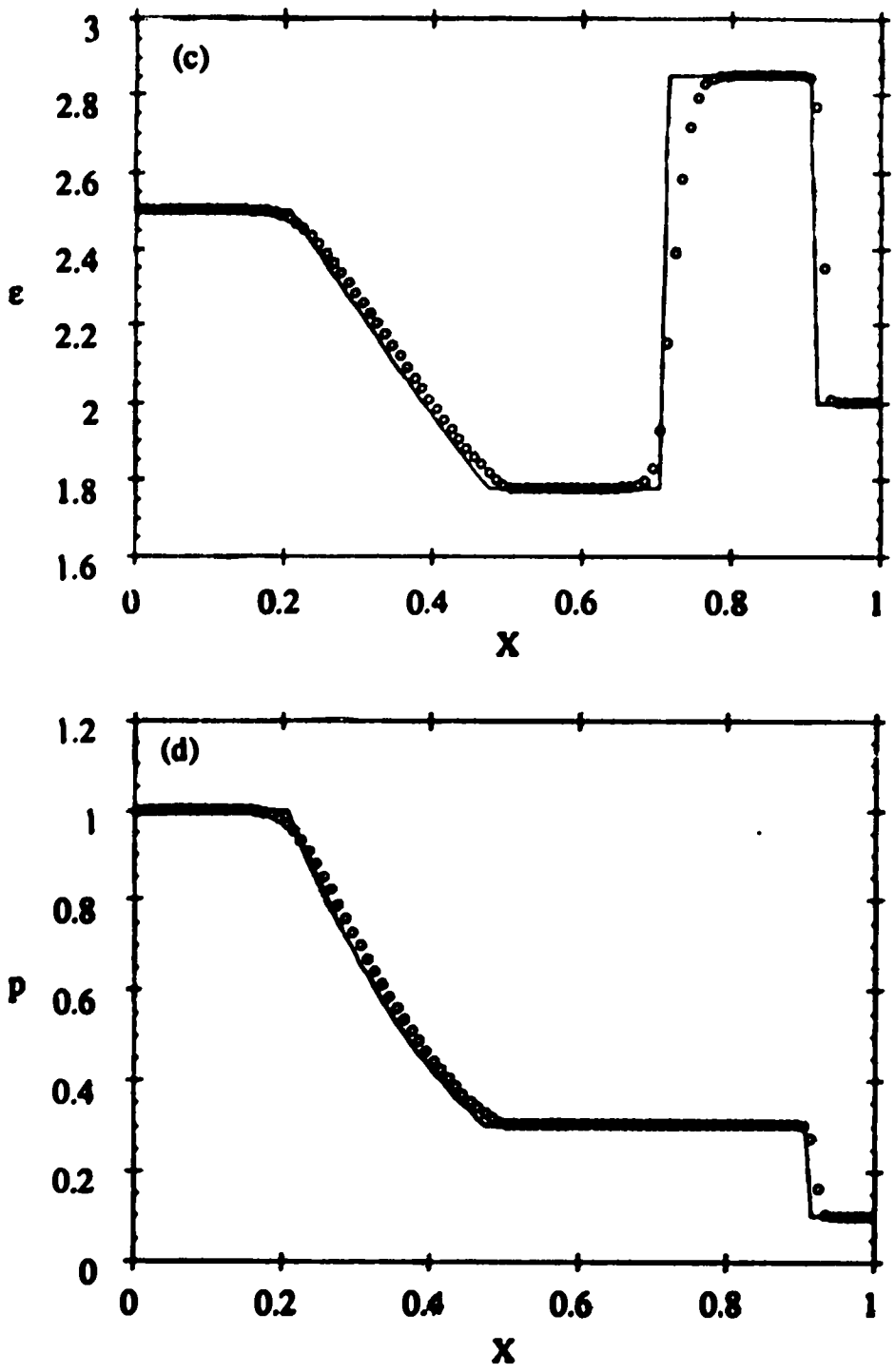


Figure 5.15: continued.

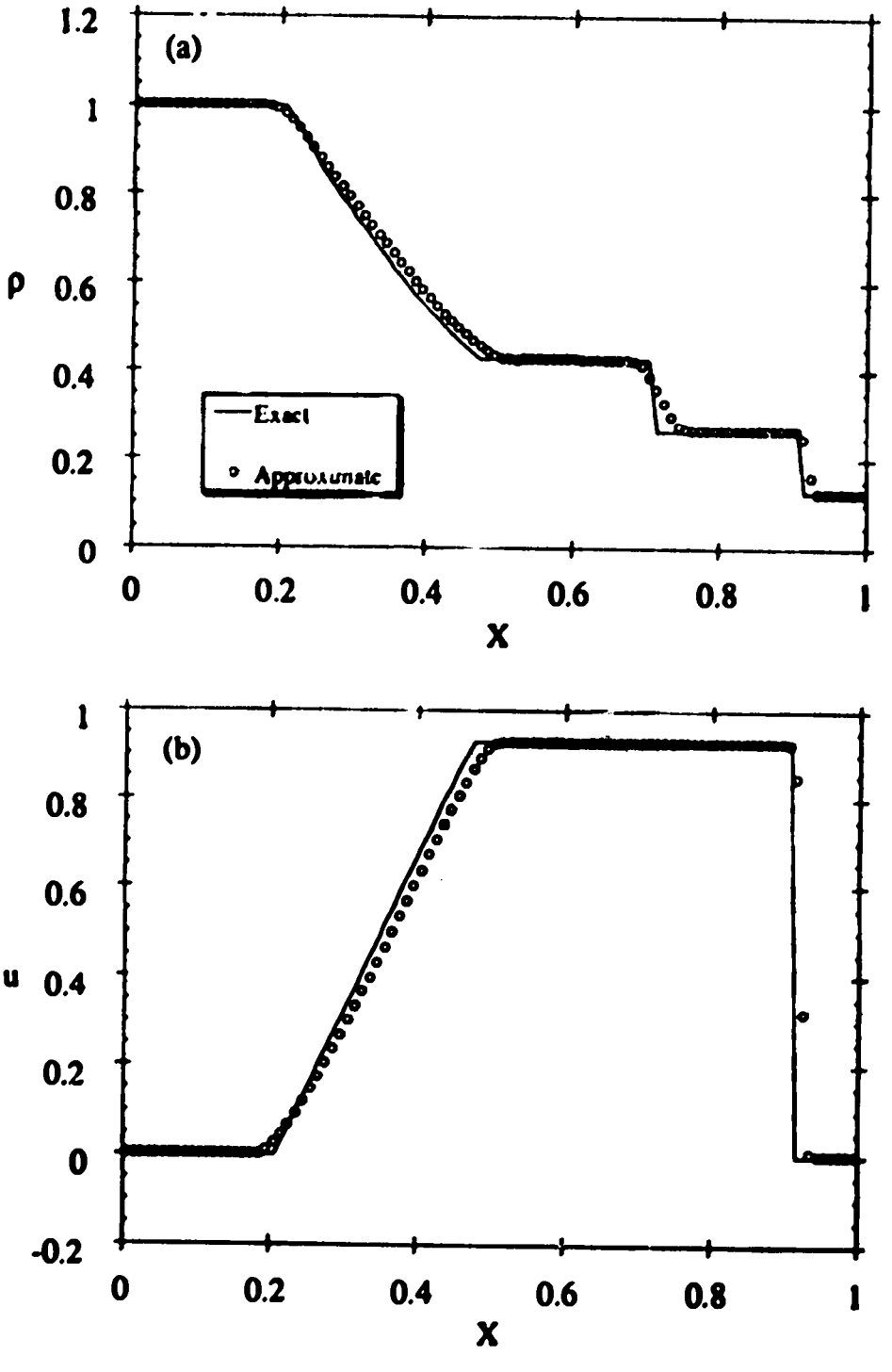


Figure 5.16: Solution of Sod's shock tube problem with the modified-flux FCT and $n = 1.5$ limiters on all fields.

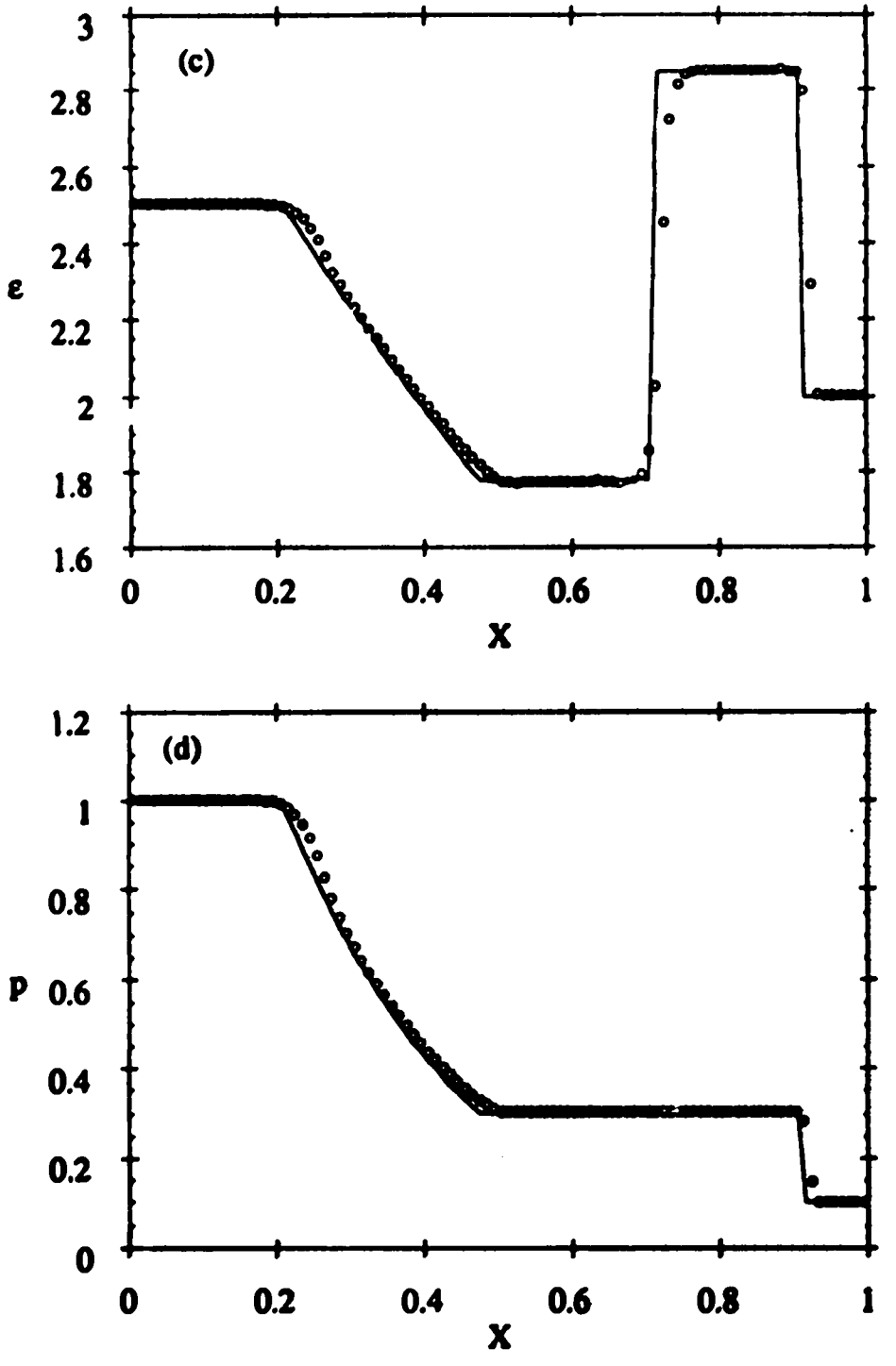


Figure 5.16: continued.

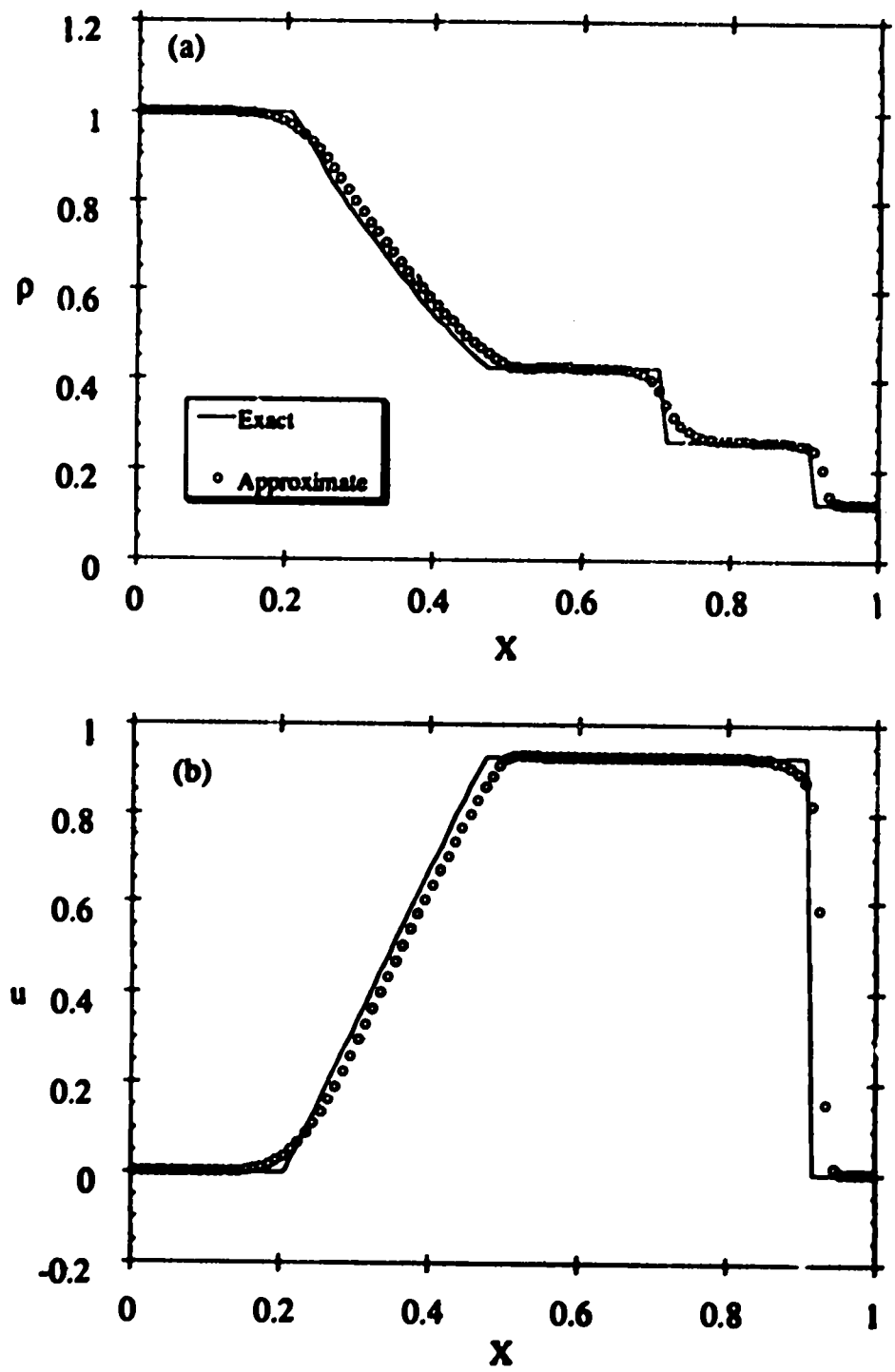


Figure 5.17: Solution of Sod's shock tube problem with a symmetric TVD algorithm.

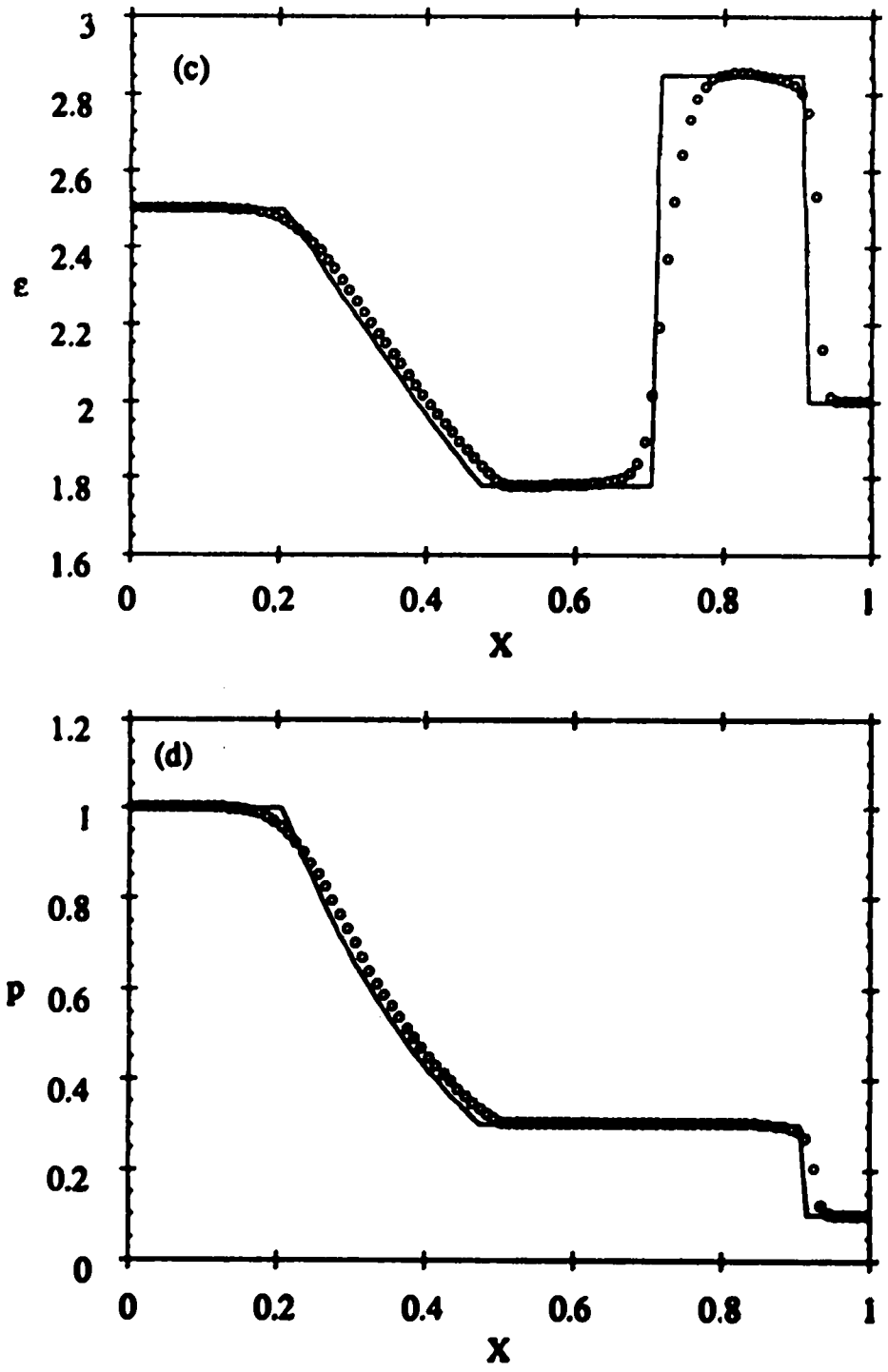


Figure 5.17: continued.

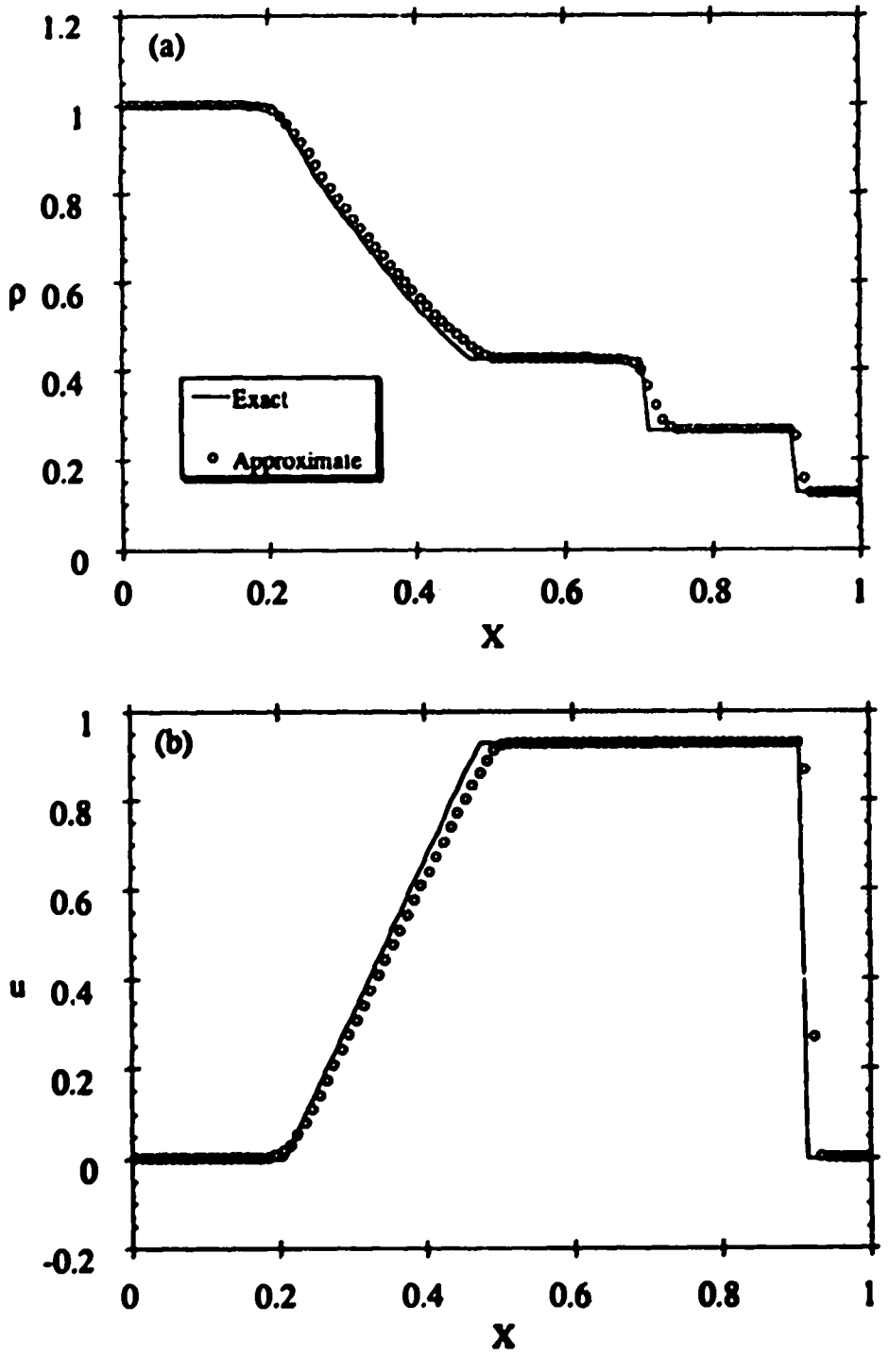


Figure 5.18: Solution of Sod's shock tube problem with a UNO limiter and a modified-flux TVD algorithm.

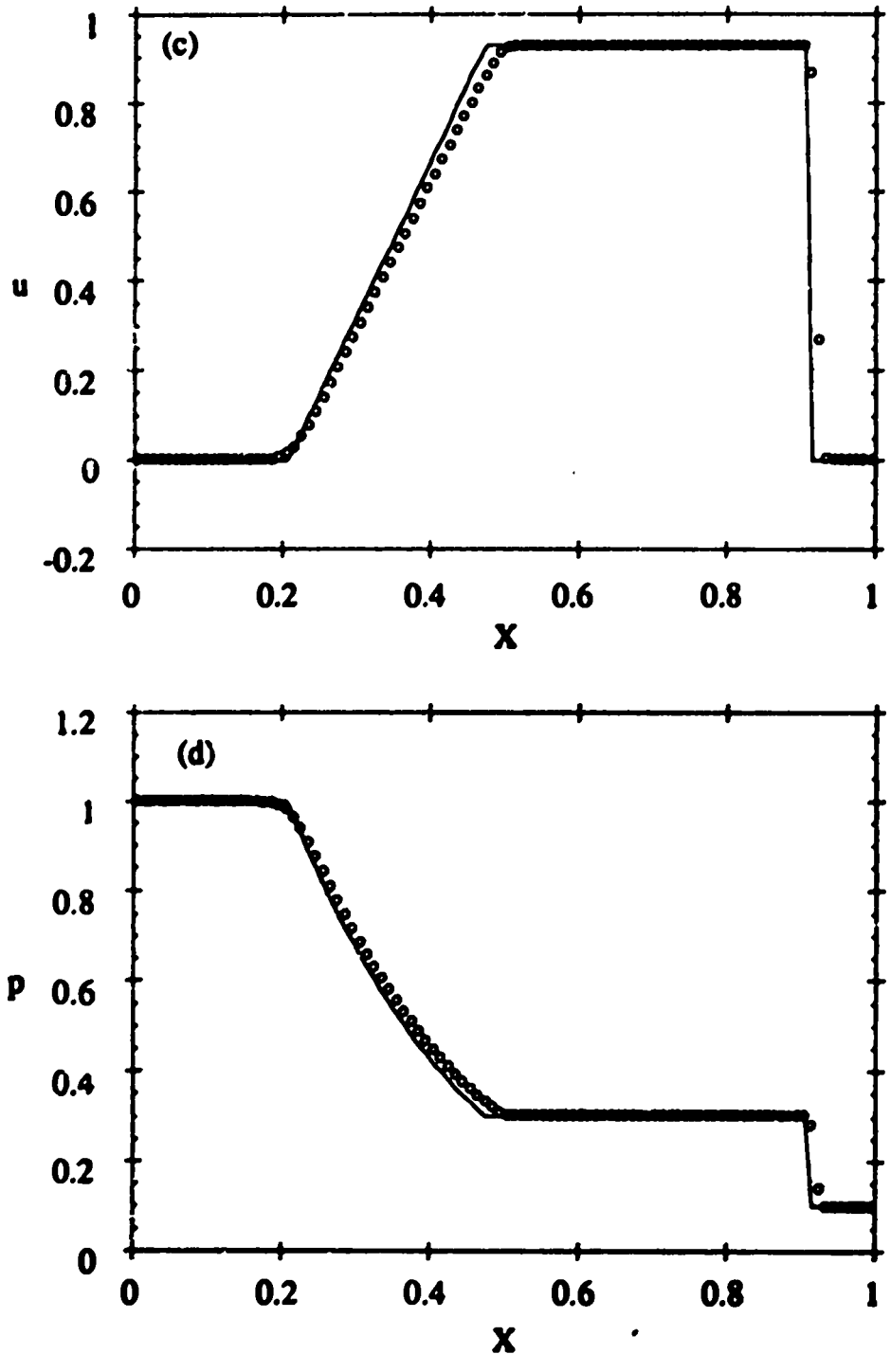


Figure 5.18: continued.

Chapter 6.

A Generalized Flux-Corrected Transport Algorithm: A Geometric Approach

---

> It is written in the language of mathematics and its characters are triangles, circles, and other geometrical figures without which it is humanly impossible to understand a single word of it; without these, one is wandering about in a dark labyrinth. *Galileo Galilei*

6.1 Introduction

The work of Godunov [56] has led to many striking advances that have been made in the numerical solution of (2.3a). In a series of papers, van Leer [120, 60] spearheaded the modern development of HOC algorithms. Godunov's method and van Leer's extensions use polynomial representations of the conserved variables in each grid cell in the process of computing the solution. These piecewise polynomials can be discontinuous at grid cell interfaces and as such require some closure at these interfaces to compute the numerical fluxes. Typically this closure uses the local solution to a Riemann problem through either an "exact" or approximate [63] Riemann solver.

Colella and Woodward [122] advanced the method developed by van Leer with their PPM. This method is still considered a premier methods for computing the solutions to (2.3a) [129]. Several theoretical advances have been made as well as the more practical ones. Harten's theory of TVD schemes [130, 61] made great strides toward understanding the theoretical properties of methods like van Leer's and those discussed below. Although these methods were first formulated as either purely Lagrangian or Eulerian through a combination of a Lagrangian step plus a remap step, these also can be used in a purely Eulerian context [123]. The methods derived in this chapter also can be used in either of these forms, but the description found below is presented in a purely Eulerian context.

Several different varieties of TVD methods have been introduced, such as the modified flux formulation from Harten and several "symmetric" TVD schemes. Roe introduced one form of TVD scheme [131]. Davis [133] also presents a method of the same general form. Sweby [132] and Roe [176] present a similar method, but the limiters are of an upwind-biased nature. Yee [134] christened these schemes as symmetric TVD schemes. The general form of symmetric TVD schemes can be looked at in several different ways: as an advanced form of artificial difusion, a Lax-Wendroff method [58] with an additional dissipative flux to ensure a TVD solution, or a TVD

method that is symmetric in its stencil whenever the limiter is not present. Another view taken in this chapter, more closely ties this formulation to that introduced by van Leer. This viewpoint has been used in the derivation of TVD methods by several authors. The TVD analog to van Leer's MUSCL scheme was discussed by Osher [179]. Goodman and LeVesque [135] took a geometric view was in deriving a TVD method.

Another modern advection algorithm also can be viewed along these lines. Perhaps the first modern algorithm to recognize the necessity of nonlinearity in the difference scheme was the method of flux-corrected transport (FCT) as introduced by Boris and Book [59]. This method was developed with the recognition of the theorem of Godunov, which states that no algorithm can be both linear and second-order accurate. This theorem does not preclude the possibility of producing a "monotone" second-order scheme, but simply states that such a method cannot be linear in nature. Thus, the FCT was a nonlinear blending of high- and low-order numerical fluxes, which ensures the lack of dispersive ripples. In a series of papers [59, 140, 141, 142, 62], this method has been revised and extended. The author recognized that the FCT and the symmetric TVD of Yee were very similar in terms of form and could easily be unified into a single general algorithm developed in Chapter 5.

At this point it is useful to delineate the difference between slope and flux limiters more closely. This is done from the standpoint of a philosophical differentiation rather than from a purely technical basis. The slope limiters can be thought of as being used directly during interpolation. Flux limiting usually involves methods that are classified as finite-difference types. Thus slope limiting applies to HOG schemes and the flux limiting applies to TVD and FCT algorithms. One caveat can be placed on this classification: it is not stringent. An example of this is the ENO schemes from Shu and Osher [65, 66], where flux limiters are used. Previous work with ENO schemes proceeded from the standpoint of slope limiters.

In extending the methods to systems of equations, the TVD and HOG type methods use Riemann solvers, which have many exceptional theoretical and aesthetic appeals. The extension of FCT, on the other hand, is usually extended in what seems an *ad hoc* formulation [143, 144]. In Lagrangian coordinates this might seem somewhat less so, as the splitting between sound waves and fluid motion is somewhat built in, but the same principles apply as with the Euler equations (see Appendix B). In this regard, I feel that there is no reason why the Riemann solvers, which have been so successful with TVD type methods, cannot be used with FCT.

With this in mind, the generalization of the FCT algorithm from a geometric point of view is discussed below. This discussion also holds for the symmetric type of TVD scheme and serve as an extension of this method. Through the use of ideas of UNO schemes, these algorithms are extended to higher than first-order accuracy in the maximum norm.

This chapter is organized into four sections. The second section first reviews modern high resolution algorithms. The geometric analog to the symmetric TVD scheme

is then introduced. This method is also extended from a linear to a quadratic reconstruction scheme. Uniformly nonoscillatory schemes are also discussed. Following this presentation, results for the schemes developed here are given for several test problems: the scalar wave equation, Burgers' equation and the Euler equations. The fourth section gives closing remarks and conclusions.

6.2 Method Development

In this section, the unified description of the symmetric TVD and FCT methods is reviewed. It should be noted that this is in a finite difference form, rather than a finite volume form. Following this brief review, the finite volume methods as typified by the Godunov and HOG algorithms are described. A tie between these methods is drawn along the same lines as the modified flux TVD scheme of Harten is related to the methods developed by van Leer. Several variants of the geometric FCT is given along with their description and mathematical properties.

6.2.1 Review of Modern Advection Algorithms

In previous work, I drew parallels between the symmetric TVD methods and the various FCT methods [6]. Specific parallels between the symmetric TVD methods and the extension of the FCT as given by Zalesak are concentrated on, with several improvements suggested for the FCT methods.

The specific form of the symmetric TVD schemes for (2.3a) is

$$u_j^{n+1} = u_j^n - \sigma \left(\hat{f}_{j+\frac{1}{2}} - \hat{f}_{j-\frac{1}{2}} \right) . \tag{6.1a}$$

where $\sigma = \Delta t / \Delta x$, $\Delta x = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$, $\Delta t = t^{n+1} - t^n$, with

$$\hat{f}_{j+\frac{1}{2}} = \frac{1}{2}(f_j + f_{j+1}) + \phi_{j+\frac{1}{2}} , \tag{6.1b}$$

being the numerical flux; also defined are $x_{j+\frac{1}{2}} = \frac{1}{2}(x_j + x_{j+1})$ and $x_{j-\frac{1}{2}} = \frac{1}{2}(x_{j-1} + x_j)$. The term $\phi_{j+\frac{1}{2}}$ is the numerical dissipation function, which is the key to obtaining high-order accuracy without dispersive ripples. For example, the form for this function for donor cell or upwind differencing is

$$\phi^{DC}_{j+\frac{1}{2}} = \frac{1}{2} \left| a_{j+\frac{1}{2}} \right| \Delta_{j+\frac{1}{2}} u , \tag{6.2}$$

where a the characteristic speed $\partial f / \partial u$, and $\Delta_{j+\frac{1}{2}} u = u_{j+1} - u_j$. If the method is used to solve a system of equations, then some modification in the definition of the above terms is in order.

For the FCT, the overall dissipation function is defined by

$$\phi^{FCT}_{j+\frac{1}{2}} = \phi^{DC}_{j+\frac{1}{2}} + \phi^{A}_{j+\frac{1}{2}} . \tag{6.3}$$

where ϕ^A is the limited difference between the a high-order flux and the donor cell flux (or another appropriate monotone scheme). This term is also known as the antidiffusive flux. The symmetric TVD scheme has its dissipation function stated as [134]

$$\phi^{SYM}_{j+\frac{1}{2}} = \left[\left(\left|a_{j+\frac{1}{2}}\right| - \sigma a^2_{j+\frac{1}{2}}\right) Q_{j+\frac{1}{2}} - \left|a_{j+\frac{1}{2}}\right|\right] \Delta_{j+\frac{1}{2}} u . \tag{6.4}$$

where $Q_{j+\frac{1}{2}}$ is a function of a the local gradients, $\Delta_{j-\frac{1}{2}} u$, $\Delta_{j+\frac{1}{2}} u$, and $\Delta_{j+\frac{3}{2}} u$ where

$$r_{j+\frac{1}{2}} = \frac{\Delta_{j+\frac{1}{2}} u}{\Delta_{j+\frac{1}{2}} x} . \tag{6.5}$$

The actual limiters used are described in detail in Chapter 8.

If the high-order flux used in the FCT is a Lax-Wendroff flux, these two methods are virtually identical. To show this requires that the flux limiter used in the FCT be changed slightly. The multipliers on the local gradient terms need to be changed from σ^{-1} to $|a| - \sigma a^2$ as suggested by the author in the previous chapter. In that chapter, parallels between both symmetric TVD and the modified flux TVD schemes and the FCT were described. The redefined FCT algorithm is shown to produce TVD results.

The modified TVD method is simply a finite difference analog to a second-order Godunov method like that of van Leer. For a scalar advection equation, the two methods are identical if the slope limiter used in the HOG method is equivalent to the flux limiter used in the TVD scheme. A HOG method is described by Algorithm 1 with the only difference being the order of the interpolation used in the reconstruction step being higher than zero. As stated earlier, this algorithm can take the form of either a totally Eulerian algorithm, or a Lagrangian solution (the local solution step) with an Eulerian remap (overall solution step). Higher order schemes are produced with higher order prescriptions (during the reconstruction step) for the function $P_j(x)$, such as those produced by MUSCL, PPM, UNO or ENO methods.

6.2.2 Geometric Symmetric TVD and FCT Schemes

The Lax-Wendroff method [58] is the canonical classical second-order method. This method produces second-order solutions, but with spurious oscillations near discontinuities, thus raising the possibility of producing negative values of positive definite values such as density or pressure. With several observations about the Lax-Wendroff method and the symmetric TVD scheme (and its relation to FCT) a geometrically based algorithm can be found. From the standpoint of algorithmic description, geometric depiction is particularly useful. Normally, the method of Lax-Wendroff is

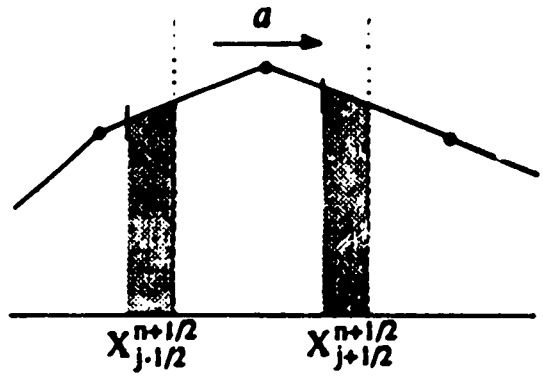


Figure 6.1: A geometric interpretation of the Lax-Wendroff method is given. This shows how this method consists of a simple linear averaging with an "upwind" correction to give time centered flux functions.

described as a finite-difference algorithm; however, it also can be described geometrically.

It is well known that the second-order central difference scheme with forward Euler time differencing is unconditionally unstable. This can be easily verified with Von Neumann stability analysis, but I proceed from a different standpoint. First, some nomenclature needs to be introduced. The flux functions for difference schemes of the form are functions of the dependent variables and can be written in terms of interpolating polynomials. Thus, given a piecewise polynomial, $P_j(x)$, that interpolates the dependent variable u, the flux functions can be written

$$f_j(u) = f[P_j(x)] \ . \tag{6.6}$$

With this definition, the problem reduces to approximating the dependent variables on a grid and computing the value of the interpolant at cell edges.

The Lax-Wendroff method was defined in Chapter 3. The symmetric TVD scheme is thought to be the Lax-Wendroff scheme plus some upwind-biased, nonlinear numerical diffusion. The canonical upwind scheme is Godunov's method, which is based on a geometric derivation. Combining this fact with the above discussion shows in a heuristic sense that the symmetric TVD scheme has a geometric analog. Now I will be somewhat more concrete in the derivation.

Lemma 1 *The symmetric TVD method can be defined in terms of the reconstructive polynomial*

$$P_j(x) = \begin{cases} u_j + s_{j+\frac{1}{2}}(x - x_j) & ; x \in \left[x_j, x_{j+\frac{1}{2}}\right] \\ u_j + s_{j-\frac{1}{2}}(x - x_j) & ; x \in \left[x_{j-\frac{1}{2}}, x_j\right] \end{cases} , \tag{6.7}$$

which is always C^1 continuous, but not C^0 continuous unless for instance $s_{j+\frac{1}{2}} = s_{j+\frac{1}{2}}$.

This requires that the cell edge slope $s_{j+\frac{1}{2}}$ be defined by some appropriate slope limiter.

Proof For the scalar advection law, $f(u) = au$, the scheme derived from the polynomial shown above can be written

$$u_j^{n+1} = u_j^n - \sigma \left\{ f\left[P_j\left(x_j^R\right), P_{j+1}\left(x_{j+1}^L\right)\right] - f\left[P_{j-1}\left(x_{j-1}^R\right), P_j\left(x_j^L\right)\right] \right\} . \tag{6.8}$$

The decision about which polynomial to use at each flux interface requires the invocation of the solution to the Riemann problem, which is simply the upwinding principle for the scalar case. Taking $a > 0$ (the case where $a < 0$ is analogous), (6.8) becomes

$$u_j^{n+1} = u_j^n - \sigma \left\{ f\left[P_j\left(x_j^R\right)\right] - f\left[P_{j-1}\left(x_{j-1}^R\right)\right] \right\} , \tag{6.9a}$$

and substituting the above definitions of x_j^L and x_j^R, (3.13a) and (3.13b), gives

$$x_j^R = x_{j+\frac{1}{2}} - \frac{a\Delta t}{2} , \quad x_{j-1}^R = x_{j-\frac{1}{2}} - \frac{a\Delta t}{2} , \tag{6.9b}$$

which in turn gives

$$P_j\left(x_j^R\right) = u_j + s_{j+\frac{1}{2}}\left(x_{j+\frac{1}{2}} - \frac{a\Delta t}{2} - x_j\right) , \tag{6.9c}$$

with P_{j-1}^R defined analogously. This equation can be simplified to

$$P_j\left(x_j^R\right) = u_j + s_{j+\frac{1}{2}}\left(\frac{\Delta x_j}{2} - \frac{a\Delta t}{2}\right) . \tag{6.9d}$$

defining $\widetilde{\Delta_{j+\frac{1}{2}}} u = s_{j+\frac{1}{2}} \Delta x_j$ and setting $\Delta x_j = \Delta x_{j-1}$. These equations can be written as

$$u_j^{n+1} = u_j^n - \sigma a \left(u_j^n - u_{j-1}^n\right) + \sigma a \left(1 - \frac{\sigma a}{2}\right)\left(\widetilde{\Delta_{j+\frac{1}{2}}} u - \widetilde{\Delta_{j-\frac{1}{2}}} u\right) . \tag{6.9e}$$

Writing the cell edge flux for the above scheme gives

$$f_{j+\frac{1}{2}} = a u_j^n + a(1 - \sigma a)\widetilde{\Delta_{j+\frac{1}{2}}} u , \tag{6.9f}$$

which can be rewritten as

$$f_{j+\frac{1}{2}} = \frac{a}{2}\left(u_j^n + u_{j+1}^n\right) - |a| \Delta_{j+\frac{1}{2}} u + \left(|a| - \sigma a^2\right)\widetilde{\Delta_{j+\frac{1}{2}}} u , \tag{6.9g}$$

where $\widetilde{\Delta_{j+\frac{1}{2}}} u$ can be written $Q_{j+\frac{1}{2}} \Delta_{j+\frac{1}{2}} u$. This is simply the symmetric TVD scheme as given by (6.1b) with (6.4) and thus is also a geometric analog to the FCT algorithm. □

In [134], the conditions for the above scheme to be TVD are stated. By writing $s_{j+\frac{1}{2}}$ as $Q(r^-, r^+) s_{j+\frac{1}{2}} = Q_{j+\frac{1}{2}}$, the conditions are modified to include the effects

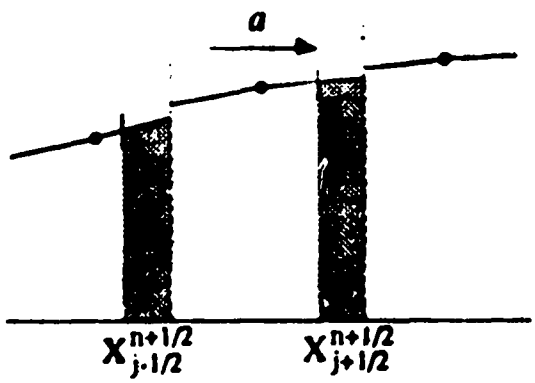


Figure 6.2: The symmetric TVD schemes geometric analog is similar to the Lax-Wendroff method, with the major difference being the limiting of the slopes. This leaves the scheme with C^1 continuity, but not C^0 continuity.

of the time centering of the fluxes ($\theta = 0$, explicit scheme using forward Euler time differencing) and are written as

$$Q_{j+\frac{1}{2}} \leq \frac{2}{1-\nu}\,, \tag{6.10a}$$

$$\frac{Q_{j+\frac{1}{2}}}{r^-} \text{ or } \frac{Q_{j+\frac{1}{2}}}{r^+} \leq \frac{2}{\nu(1-\nu)} - \frac{2}{1-\nu}\,, \tag{6.10b}$$

and

$$\nu \leq 1\,. \tag{6.10c}$$

This assumes that both Q and Q/r are positive. Without these assumptions the conditions above take a more complicated form, but allow a slightly larger set of Q functions.

Figure 6.2 shows the pictorial representation of this scheme. For the scalar wave equation, this method and the classic symmetric TVD are equivalent, but for nonlinear problems the two methods are as different as Harten's modified TVD is different from the corresponding MUSCL scheme.

6.2.3 Parabolic Symmetric TVD and FCT Schemes

If one proceeds along this line of thought and considers a polynomial approximation, it is notable that three conditions exist for each grid cell in the above scheme, and that one degree of freedom is not fully utilized. These conditions are

$$P_j(x_j) = u_j\,;\; \frac{dP_j}{dx}\left(x_{j-\frac{1}{2}}\right) = s_{j-\frac{1}{2}}\,;\; \frac{dP_j}{dx}\left(x_{j+\frac{1}{2}}\right) = s_{j+\frac{1}{2}}\,.$$

thus a unique parabola can be fit in each cell. Taking the form

$$P_j(\theta) = A_j(x - x_j)^2 + B_j(x - x_j) + C_j \,. \tag{6.11a}$$

the coefficients are defined

$$A_j = \frac{\dot{s}_{j+\frac{1}{2}} - \dot{s}_{j-\frac{1}{2}}}{2\Delta x_j} \,. \tag{6.11b}$$

$$B_j = \frac{\dot{s}_{j+\frac{1}{2}} + \dot{s}_{j-\frac{1}{2}}}{2} \,. \tag{6.11c}$$

and

$$C_j = u_j \,. \tag{6.11d}$$

Thus, the interpolant can be written for completeness:

$$P_j(\theta) = u_j + \left(\frac{\dot{s}_{j-\frac{1}{2}} + \dot{s}_{j+\frac{1}{2}}}{2}\right)(x - x_j) + \left(\frac{\dot{s}_{j+\frac{1}{2}} - \dot{s}_{j-\frac{1}{2}}}{2\Delta x_j}\right)(x - x_j)^2 \,.$$

This polynomial describes what I call the parabolic FCT when used with the convective algorithm described by (6.8). It should be noted that the temporal integration can be accomplished by other means such as a multistage algorithm.

I now seek to prove under what conditions this algorithm produces TVD results. These conditions define the allowable values of the cell edge slopes, $\dot{s}_{j\pm\frac{1}{2}}$.

Theorem 6 *The parabolic symmetric TVD and FCT method derived above is TVD under the following conditions:*

1. *If the slopes $\dot{s}_{j\pm\frac{1}{2}}$ can be of opposite sign, the function $Q(r^-, 1, r^+)$ must be less than or equal to $|4/3|$.*

2. *If the slopes $\dot{s}_{j\pm\frac{1}{2}}$ are required to be of the same sign, the function $Q(r^-, 1, r^+)$ must be less than or equal to $8/3$.*

Proof. For the following proof, only the spatially accurate case is studied, thus to some extent this study is limited to the semi-discrete version of the equation. Thus the TVD conditions [180] shown above are simplified to

$$\frac{\partial u}{\partial t} = C_j \Delta_{j+\frac{1}{2}} u - D_j \Delta_{j-\frac{1}{2}} u \,. \tag{6.12a}$$

$$C_j, \; D_j \geq 0 \,. \tag{6.12b}$$

For time integration typically a Lax-Wendroff or Cauchy-Kowsaleski procedure is applied, which in some sense is characteristic tracing. Runge-Kutta algorithms also can be used, although for the corresponding composite algorithm, the Runge-Kutta

methods are not classical in form [160]. In general, careful analysis must be applied to determine the stability requirements.

Examining the case where $a > 0$, with the case where $a < 0$ yielding equivalent results. Given this characteristic speed, (6.12a) with (6.11a) becomes

$$\frac{\partial u}{\partial t} = -\frac{a}{\Delta x}\left(u_j^n - u_{j-1}^n\right) - a\left[\left(\frac{3}{8}\dot{s}_{j+\frac{1}{2}} + \frac{1}{8}\dot{s}_{j-\frac{1}{2}}\right) - \left(\frac{3}{8}\dot{s}_{j-\frac{1}{2}} + \frac{1}{8}\dot{s}_{j-\frac{3}{2}}\right)\right] . \tag{6.13}$$

Setting $C_j = 0$ and rewriting the above equation in a form amenable to analysis produces

$$\frac{\partial u}{\partial t} = -a\left[1 + \left(\frac{3}{8}\frac{Q_{j+\frac{1}{2}}}{r^-} - \frac{1}{4}Q_{j-\frac{1}{2}} - \frac{1}{8}\frac{Q_{j-\frac{3}{2}}}{r^+}\right)\right] s_{j-\frac{1}{2}} . \tag{6.14}$$

It should be noted that all the three parameter limiters that would be used with the above formulation are a function of $\Delta_{j-\frac{1}{2}}u$, and the Q limiters are function conservative gradients [132, 176]. Putting this form into the form useful for analysis and using the TVD conditions discussed above

$$\frac{a}{\Delta x}\left[1 + \left(\frac{3}{8}\frac{Q_{j+\frac{1}{2}}}{r^-} - \frac{1}{4}Q_{j-\frac{1}{2}} - \frac{1}{8}\frac{Q_{j-\frac{3}{2}}}{r^+}\right)\right] \geq 0 , \tag{6.15}$$

allows the proper conditions on $Q(u)$ to be established for TVD solutions. If I set $Q_{j-\frac{3}{2}}/r^+ = Q_{j-\frac{1}{2}}$ as a bound and simplify accordingly, the above condition becomes

$$a\left[1 + \left(\frac{3}{8}\frac{Q_{j+\frac{1}{2}}}{r^-} - \frac{3}{8}Q_{j-\frac{1}{2}}\right)\right] \geq 0 . \tag{6.16}$$

This simplification seems a quite reasonable bound in light of the functional form of the flux/slope limiters.

For the first of the two cases, the proof is

$$\frac{3}{8}\left(Q_{j-\frac{1}{2}} - \frac{Q_{j+\frac{1}{2}}}{r^-}\right) \leq 1 , \tag{6.17}$$

which gives the condition that $|Q(u)| \leq 8/3$. This corresponds to the limiter of the "minbar" type that is defined by

$$\bar{m}_a = \begin{cases} a\,a & |a| = \inf\left(|a|, |b|, |c|\right) \\ a\,b & |b| = \inf\left(|a|, |b|, |c|\right) \\ a\,c & \text{otherwise} \end{cases} . \tag{6.18}$$

where a is a constant that is $0 \leq a \leq 4/3$ to produce a TVD solution.

Before going onto the second case, certain caveats should be applied to this class

of limiter. Although the "minbar" limiter is a TVD limiter in the sense of Harten's definition of TVD schemes, it is not a classic "monotonicity" limiter, similar to the type derived by van Leer [120, 60], and thus has some fewer favorable geometric properties. The act of not necessarily clipping at extrema yields construction of new extrema near extrema, in the data, which are not necessarily physical. This may not be much of a problem if one takes the ENO philosophy of simply seeking the smoothest available interpolant within some local support. Nevertheless, care should be taken in applying this limiter as the results section shows.

The second case proceeds much in the same way and yields a class of limiters that are very similar to the "classic" TVD limiters. For the above-stated conditions for positive definite values of $Q(u)$ changes the form of (6.17) to

$$\frac{3}{8}Q_{j-\frac{1}{2}} \leq 1 \tag{6.19a}$$

and

$$\frac{3}{8}Q_{j+\frac{1}{2}} \leq r^- \,. \tag{6.19b}$$

which gives a limiter such that $0 \leq Q(u) \leq 8/3$. In the same fashion as TVD limiters, the compression applied by the limiter grows with the increasing value of the limiter maximum. Thus the limiter associated with the scalar, 8/3, would correspond to the "superbee" limiter defined by Roe [176]. □

A three-parameter limiters of the form discussed earlier are within this class. In addition, some general useful forms of this class of limiter would be

$$Q_{j\pm\frac{1}{2}} = \mathrm{m}\left[\frac{4}{3}r^-, \frac{4}{3}, \frac{4}{3}r^+, \frac{1}{2}\left(r^- + r^+\right)\right] \tag{6.20a}$$

and

$$Q_{j\pm\frac{1}{2}} = \mathrm{m}\left[\frac{8}{3}r^-, \frac{8}{3}, \frac{8}{3}r^+, \frac{1}{2}\left(r^- + r^+\right)\right] \,. \tag{6.20b}$$

The order of accuracy of the limiters discussed above provides the parabolic FCT algorithm. To do this, the methods described by Sweby [132] will be used. Without difficulty it can be shown that the same region of the limiter curves can be obtained if the limiters discussed by Sweby are multiplied by 4/3.

A problem with this method common to all typical second-order (or higher) TVD methods is that they are order one accurate in the L_∞ norm [64]. To overcome this requires that the method be reformulated.

Using the upwind, two parameter limiters in conjunction with this method would violate the assumption made is simplifying (6.15) to (6.16). From a heuristic standpoint, this would imply the use of data at points downwind of the limiter's stencil, which would lead to instabilities.

6.2.4 UNO Symmetric TVD and FCT Schemes

To give the method described in the previous section, higher than first-order accuracy in the l_∞ norm, the symmetric and parabolic schemes are redefined by changing the form of the slope limiters.

The following lemma motivates the first of these proposed schemes:

Lemma 2 *The interpolant defined by (6.7) interpolating in the interval* $\left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$ *has a local maximum or minimum in this interval if and only is the slopes,* $\dot{s}_{j-\frac{1}{2}}$ *and* $\dot{s}_{j+\frac{1}{2}}$ *are opposite in sign.*

Proof. To prove this, take the derivative of the polynomial defined by (6.7) giving

$$\frac{dP(x)}{dx} = \begin{cases} \dot{s}_{j-\frac{1}{2}} & x \in \left[x_{j-\frac{1}{2}}, x_j\right] \\ \dot{s}_{j+\frac{1}{2}} & x \in \left[x_j, x_{j+\frac{1}{2}}\right] \end{cases} . \tag{6.21}$$

A monotone piecewise interpolant has the same sign across the interval it interpolates. If the derivative changes sign in the interval, an extrema exists in that interval. Simple inspection indicates that to produce an interpolant with a extrema requires that the cell-edged slopes differ in sign. This shows that $\dot{s}_{j+\frac{1}{2}}\dot{s}_{j-\frac{1}{2}} < 0$ produces an extrema in the local interpolant. □

Corollary 1 (Lemma 2) *If the slopes defining (6.7) are of the same sign, the interpolant is monotone on the interval* $\left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$.

Proof. To state that the interpolant is not monotone on this interval would contradict Lemma 2 and the definition of monotone interpolation (in a local sense). □

Lemma 3 *The parabola defined by (6.11a)-(6.11d) interpolating in the interval* $\left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$ *has a local maximum or minimum in this interval if and only if the slopes are opposite in sign.*

Proof. To prove this take the derivative of the polynomial defined by (6.11a) giving

$$\frac{dP(x)}{dx} = \left(\frac{\dot{s}_{j+\frac{1}{2}} - \dot{s}_{j-\frac{1}{2}}}{\Delta x}\right)(x - x_j) + \frac{\dot{s}_{j+\frac{1}{2}} + \dot{s}_{j-\frac{1}{2}}}{2} . \tag{6.22a}$$

By setting the derivative to zero the local minima and maxima can be found by

$$x^* = x_j + \frac{\Delta x\left(\dot{s}_{j+\frac{1}{2}} + \dot{s}_{j-\frac{1}{2}}\right)}{2\left(\dot{s}_{j-\frac{1}{2}} - \dot{s}_{j+\frac{1}{2}}\right)} .$$

By setting the conditions for a local extrema to lie in the interval

$$x_{j+\frac{1}{2}} \le x^* \le x_{j+\frac{1}{2}} . \tag{6.22b}$$

The values for the slopes that satisfy this inequality can be found through substitution giving

$$\tilde{s}_{j-\frac{1}{2}} \geq 0 \,, \; \tilde{s}_{j+\frac{1}{2}} \leq 0 \,, \tag{6.22c}$$

and by using symmetry this implies that

$$\tilde{s}_{j-\frac{1}{2}} \leq 0 \text{ and } \tilde{s}_{j+\frac{1}{2}} \geq 0 \tag{6.22d}$$

also satisfies the inequalities. As with Lemma 2, this shows that $\tilde{s}_{j+\frac{1}{2}}\tilde{s}_{j-\frac{1}{2}} < 0$ produces an extrema in the local interpolant. In addition, this inequality shows that if the signs of the slopes are the same any local extrema, lies outside the interpolated interval. □

Corollary 2 (Lemma 3) *If the slopes defining (6.11a) are of the same sign, the parabola is monotone in the interval* $\left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$.

Proof. To state that the interpolant is not monotone in this interval would be a contradiction of Lemma 3 and the definition of monotone interpolation (in a local sense). □

This might cause one to assume that the minbar limiter would suffice here to provide the correct slopes near minima or maxima in the data. But, one problem is that the three parameter form of the minbar limiter also would allow extrema to be found in cells where no such extrema exists in the data (to the left or the right of a true extrema).

Definition 4 (Harten and Osher [136]) *Non-oscillatory interpolation is defined by interpolation* $P_j(x)$ *that has its number of extrema in an interval that is not exceeded by the local extrema in the data,* $u(x)$.

An UNO type scheme can be derived by considering a formulation that is close the original UNO scheme. These schemes are also motivated by the desire to have a better grasp on higher order accuracy with the parabolic formulation. I begin by defining second-order accurate candidate slopes for the limiters. Consider the determination of $\tilde{s}_{j+\frac{1}{2}}$, which requires candidate slopes $s_{j-\frac{1}{2}}$, $s_{j+\frac{1}{2}}$ and $s_{j+\frac{3}{2}}$. The candidate slope $s_{j+\frac{1}{2}}$ is already second order in its standard form,

$$s_{j+\frac{1}{2}} = \frac{u_{j+1} - u_j}{x_{j+1} - x_j} \,, \tag{6.23a}$$

because it is a centered approximation about $x_{j+\frac{1}{2}}$, but the other slopes are not. In order to make these approximations second-order at $x_{j+\frac{1}{2}}$, a corrective term is needed. By expanding the definition of $s_{j+\frac{1}{2}}$ in a Taylor series about $x_{j-\frac{1}{2}}$ and $x_{j+\frac{3}{2}}$,

the following approximations are found:

$$s_{j+\frac{1}{2}} = s_{j-\frac{1}{2}} + \Delta x_j \left.\frac{ds}{dx}\right|_{x_{j-\frac{1}{2}}} + \mathcal{O}\left(\Delta x_j^2\right) , \tag{6.23b}$$

and

$$s_{j+\frac{1}{2}} = s_{j+\frac{3}{2}} - \Delta x_{j+1} \left.\frac{ds}{dx}\right|_{x_{j+\frac{1}{2}}} + \mathcal{O}\left(\Delta x_{j+1}^2\right) , \tag{6.23c}$$

where $\Delta x_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$.

Theorem 7 *The method for polynomial reconstruction described by (6.7) or (6.11a) are uniformly non-oscillatory by Definition 4 if the cell edge slopes are prescribed as follows:*

$$\dot{s}_{j-\frac{1}{2}} = m\left(s_{j-\frac{1}{2}} + s'_{j-\frac{1}{2}}\Delta x_{j-1}, s_{j-\frac{1}{2}}, s_{j+\frac{1}{2}} - s'_{j+\frac{1}{2}}\Delta x_j\right) , \tag{6.24a}$$

and

$$\dot{s}_{j+\frac{1}{2}} = m\left(s_{j-\frac{1}{2}} + s'_{j-\frac{1}{2}}\Delta x_j, s_{j+\frac{1}{2}}, s_{j+\frac{3}{2}} - s'_{j+\frac{3}{2}}\Delta x_{j+1}\right) , \tag{6.24b}$$

where $s' = ds/dx$ is defined in a consistent fashion.

Proof. For this proof, as before, I must show that the extrema in the polynomial, $P_j(x)$, coincide with the extrema in the given data. As stated in Lemmas 2 and 3, an extrema can only occur if $\dot{s}_{j-\frac{1}{2}}\dot{s}_{j+\frac{1}{2}} < 0$. A condition in the data of $s_{j-\frac{1}{2}}s_{j+\frac{1}{2}} < 0$ also signals the presence of an extrema in the data.

The consistent forms for s' considered here are

$$s'_{j+\frac{1}{2}} = \mathrm{m}\left(\frac{s_{j+\frac{3}{2}} - s_{j+\frac{1}{2}}}{\Delta x}, \frac{s_{j+\frac{1}{2}} - s_{j-\frac{1}{2}}}{\Delta x}\right) , \text{ or } \mathrm{m}\left(\frac{s_{j+\frac{3}{2}} - s_{j+\frac{1}{2}}}{\Delta x}, \frac{s_{j+\frac{1}{2}} - s_{j-\frac{1}{2}}}{\Delta x}\right) , \tag{6.25a}$$

with a similar function for $s'_{j-\frac{1}{2}}$, $s'_{j-\frac{3}{2}}$ and $s'_{j+\frac{3}{2}}$. The limited slope functions (6.24a) and (6.24b) can be written in a form similar to the Q functions introduced earlier:

$$\dot{s}_{j-\frac{1}{2}} = \mathrm{m}\left(r^- + \frac{s'_{j-\frac{1}{2}}}{s_{j-\frac{1}{2}}}\Delta x_{j-1}, 1, r^+ - \frac{s'_{j+\frac{1}{2}}}{s_{j-\frac{1}{2}}}\Delta x_j\right) s_{j-\frac{1}{2}} , \tag{6.25b}$$

and

$$\dot{s}_{j+\frac{1}{2}} = \mathrm{m}\left(r^- + \frac{s'_{j-\frac{1}{2}}}{s_{j+\frac{1}{2}}}\Delta x_j, 1, r^+ - \frac{s'_{j+\frac{3}{2}}}{s_{j+\frac{1}{2}}}\Delta x_{j+1}\right) s_{j+\frac{1}{2}} , \tag{6.25c}$$

These functions take on the same sign as $s_{j-\frac{1}{2}}$ and $s_{j+\frac{1}{2}}$, respectively, by the definition of the minmod limiter. Thus an extrema in the interpolant exists in the interval only if the extrema exists in the data by Lemmas 2 and 3. □

Remark 21 *Each of the methods discussed above can be used as an implicit algorithm. The theory surrounding the TVD methods [130, 61] gives a firm basis for*

implicit solutions and this basis follows to the application of the methods presented here.

6.3 Results

The results section of this chapter shows the strengths and weaknesses of the algorithms described above. The scalar wave equation should reveal the basic properties of the solution schemes in a simple setting. These properties hold with the use of the method in more complicated situations. Burgers' equation provides results for a nonlinear equation as well as convergence results, which show the order of accuracy obtained by the method. Finally, the Euler equations provide an indication of these algorithms performance with problems with systems of equations. For the remainder of the discussion, the following nomenclature is used:

- the standard geometric analog to the symmetric TVD scheme is denoted by the name symmetric,
- the parabolic variant of this method is denoted by quadratic
- the UNO modification of the symmetric method is denoted as the symmetric UNO, and
- the UNO modification of the quadratic method is denoted as the quadratic UNO.

A detailed account of the test problems used is given in Appendix A. Specific details of their use is given below.

6.3.1 Scalar Wave Equation

To begin to assess the algorithms presented here, a simple standard test problem was solved. On a domain of 100 equidistantly spaced cells, a square wave 10 cells in width is advected at a unit velocity with periodic boundary conditions. The CFL number is held at $\frac{1}{2}$ and the solution proceeds for 300 time steps.

The symmetric scheme performs with the lowest resolution of the schemes discussed here and has some symmetry problems as shown in Fig. 6.3. This sort on unsymmetrical behavior was noted by Munz [181] in a study of solutions to two-dimensional problems by high-resolution methods. This lack of symmetry is somewhat alleviated by the use of the quadratic scheme (see Fig. 6.4). The UNO-type methods both give significantly better solutions in terms of preservation of maximum values, but also give rise to some controlled oscillations (see Figs. 6.5 and 6.6). The quadratic method provides both better resolution than the symmetric scheme and also shows much better solution symmetry. Part of this increase in resolution can be

Table 6.1: Order of accuracy in several norms for the schemes solving Burgers' equation when the solution is smooth.

Scheme	L_1	L_2	L_∞
Symmetric	1.83	1.58	1.19
Quadratic	1.88	1.61	1.25
Symmetric UNO	1.94	1.65	1.07
Quadratic UNO	1.97	1.60	1.02

attributed to the more compressive form of the limiter used with this method ($Q_{4/3}$ rather than Q_1 and $Q_{8/3}$ rather than Q_2). When the same limiter is used in each scheme, the solution is only slightly better with the quadratic scheme; however, the quality of the results remains improved with respect to symmetry.

6.3.2 Burgers' Equation

The solution of Burgers' equation using these methods can provide more information concerning the behavior of the algorithms. By computing the error as compared with the exact solution an order of accuracy can be obtained.

When the solution is smooth, each of the solution methods is well behaved and gives convergence at expected rates as shown in Table 6.1. The UNO solutions are the most accurate and have the lowest error as well as the highest rates of convergence (especially in the L_2 norm). When a shock has formed, this situation changes in several respects. All the methods converge more slowly, but the UNO schemes converge more slowly than the simpler symmetric and quadratic schemes (see Table 6.2). The L_∞ norm also shows a "knee" in each case. This signals a slowing in the rate of convergence beyond a certain grid spacing. These results are summarized by Figs. 6.7-6.10.

For times after $t = 1.0$ the UNO solutions resume their initially high rates of convergence. The behavior shown near $t = 1.0$ seems to be temporary and limited to a short period near the formation of the shock. The poorer convergence may be related to the width of the finite difference stencil used in these schemes. This behavior was noted in [6] and was noticeable for schemes with three rather than two parameter limiters. The effect of the three parameter limiters is to increase the support of the interpolation at each cell edge. This increase is not accompanied by a subsequent increase in accuracy and because a minimum principle is used with the limiters, the effect is to lower order of accuracy due to the limiter over a wider set of grid points.

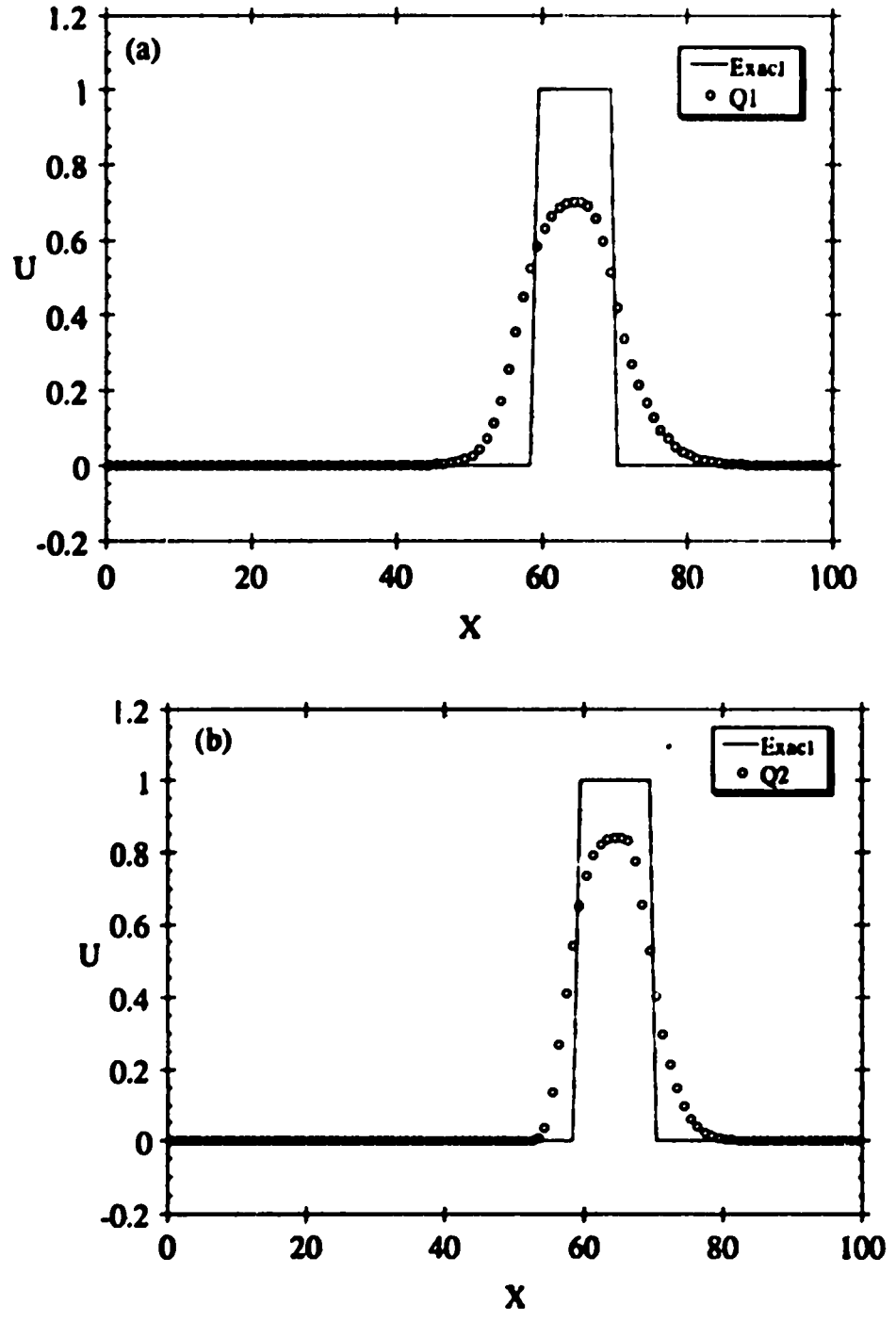


Figure 6.3: The solution of the scalar wave equation by the symmetric method using both a noncompressive, Q_1, and compressive limiter, Q_2. The Q_1 (6.3a) limiter produces a solution which is significantly better than a first-order upwind solution, but exhibits excessive smearing from diffusion. The compressive limiter (6.3b) shows an improvement in the solution as a result of reduced diffusion. Both solutions exhibit some lack of symmetry which is indicative of this method.

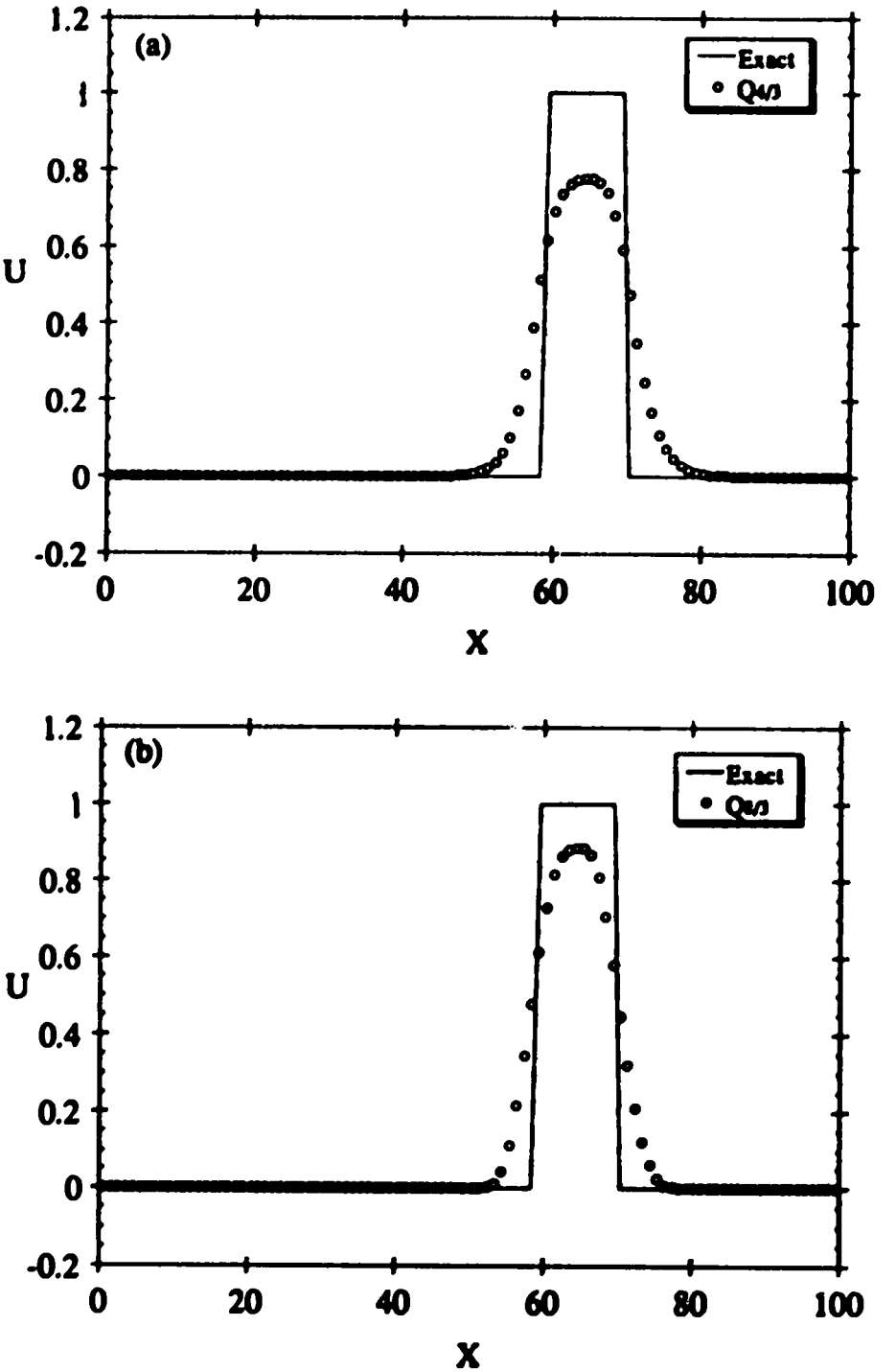


Figure 6.4: The solution of the scalar wave equation by the quadratic method using both a noncompressive, $Q_{4/3}$, and compressive limiter, $Q_{8/3}$. Again, the noncompressive limiter produces a solution that is diffused by comparison to the solution found with the compressive limiter (6.4b). Both solutions have improved symmetry when compared with the symmetric method.

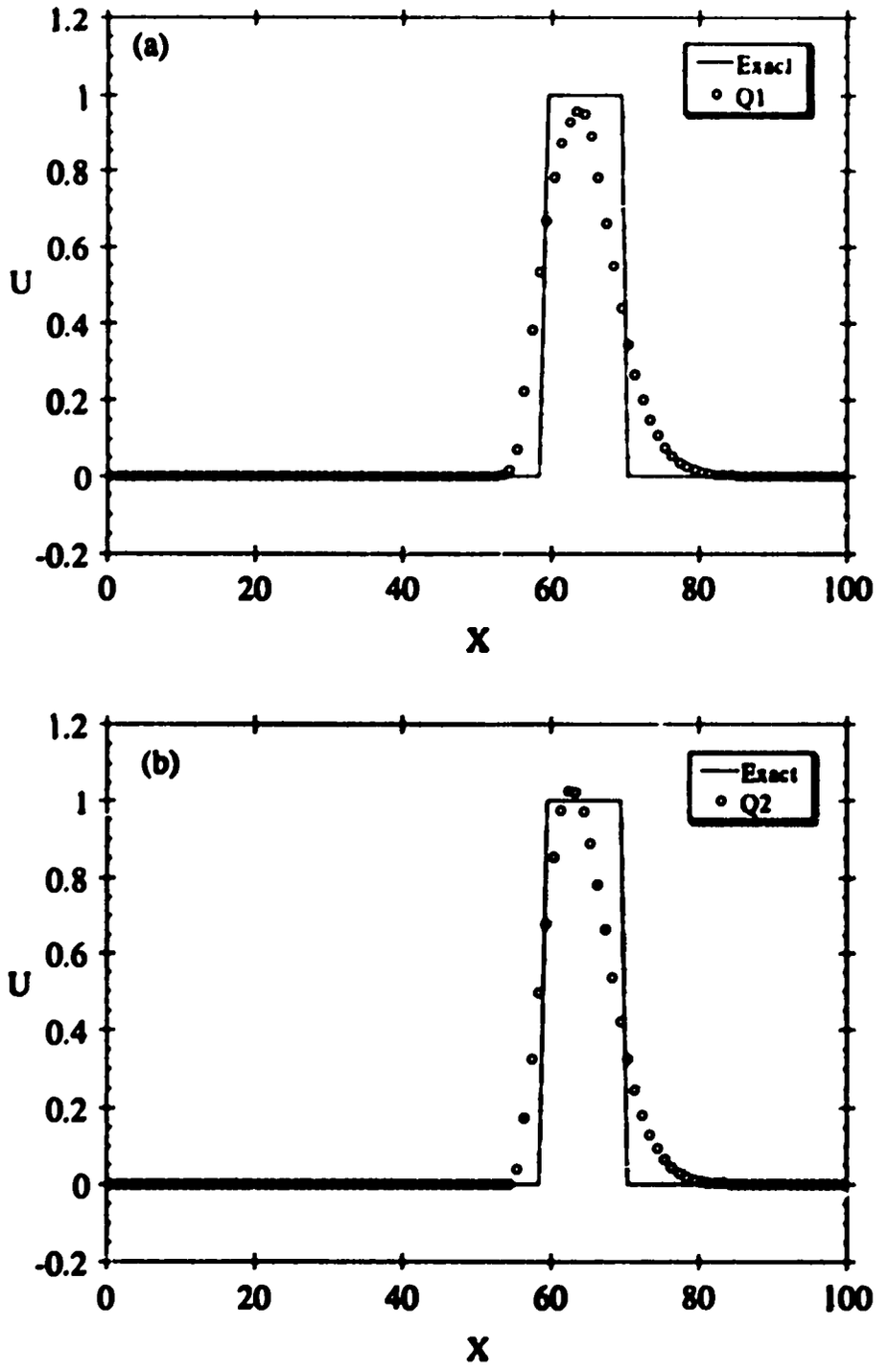


Figure 6.5: The symmetric UNO solution shows a marked increase in the preservation of the maximum value; however, the effects of a lack of symmetry are also evident. Both solutions exhibit a leading phase error greater than that present with the symmetric scheme.

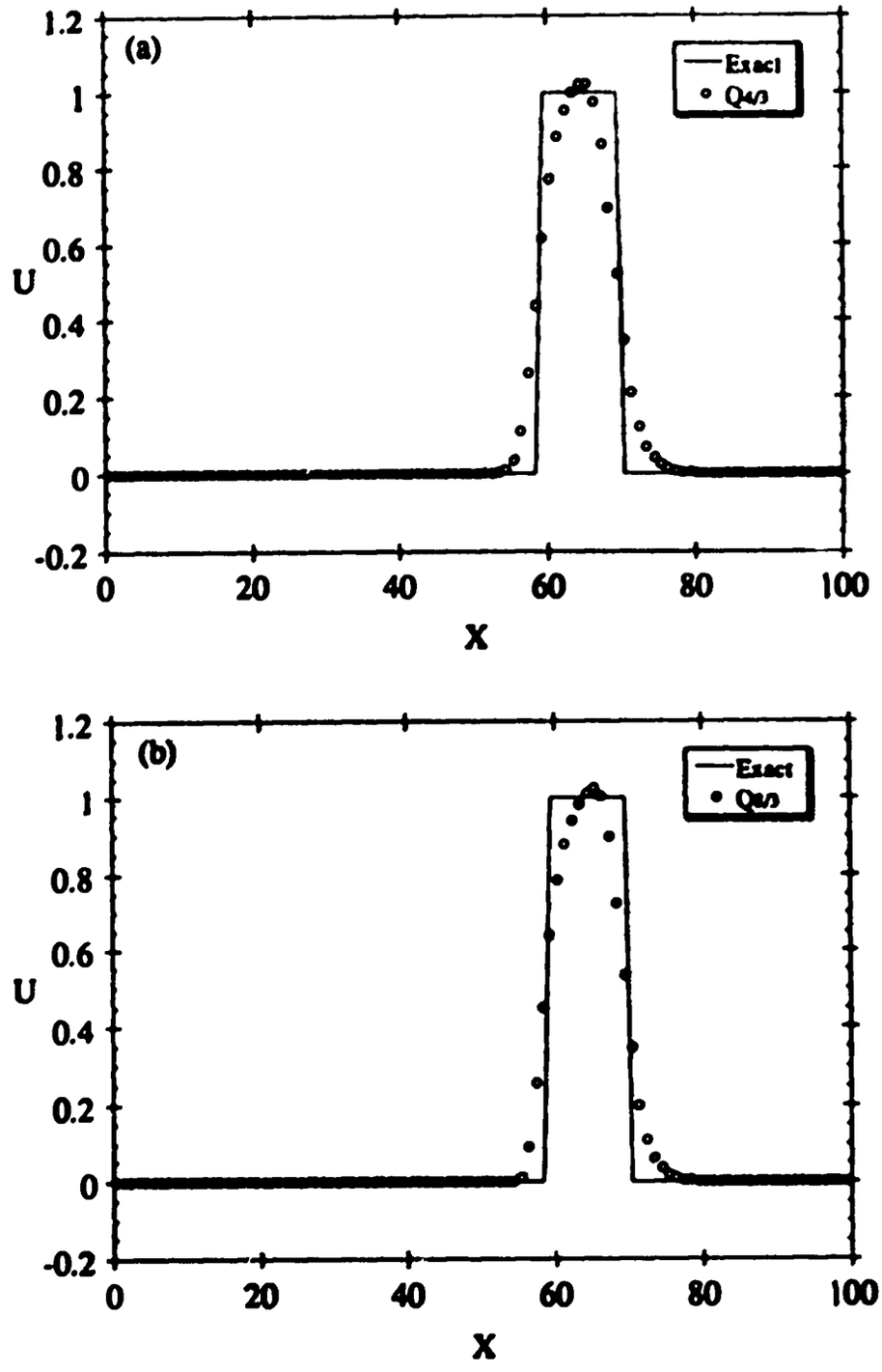


Figure 6.6: The quadratic UNO scheme gives maximum values slightly greater than the maximum value of the initial distribution. The leading phase error present in the symmetric scheme is improved somewhat. The compressive limiter gives the least additional resolution in this case.

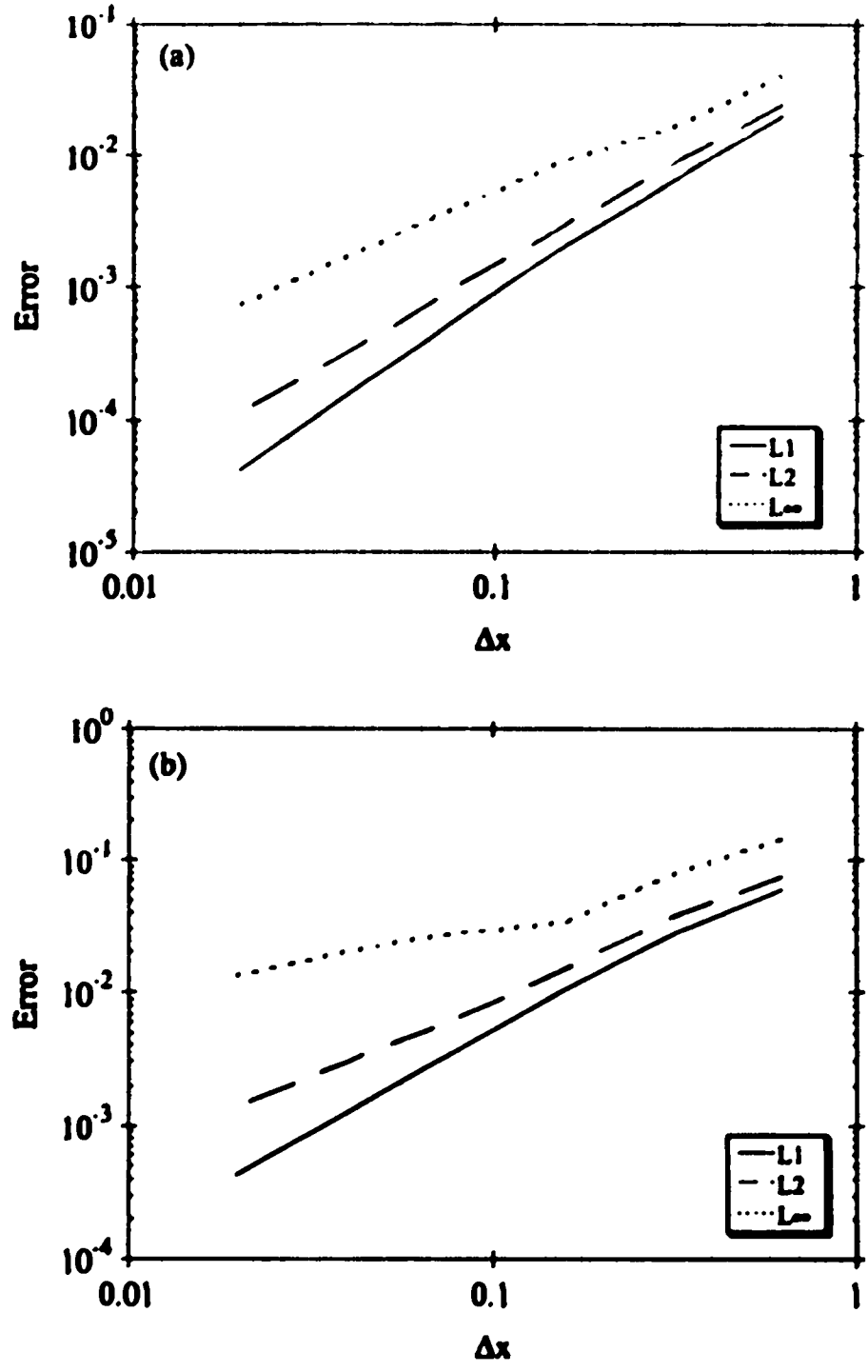


Figure 6.7: The symmetric scheme gives good, well-behaved convergence when the solution is smooth ($t = 0.2$), but when a shock forms ($t = 1.0$), the error grows by about an order of magnitude and the L_∞ norm's curve has a "knee" in it indicating a reduction in the order of convergence.

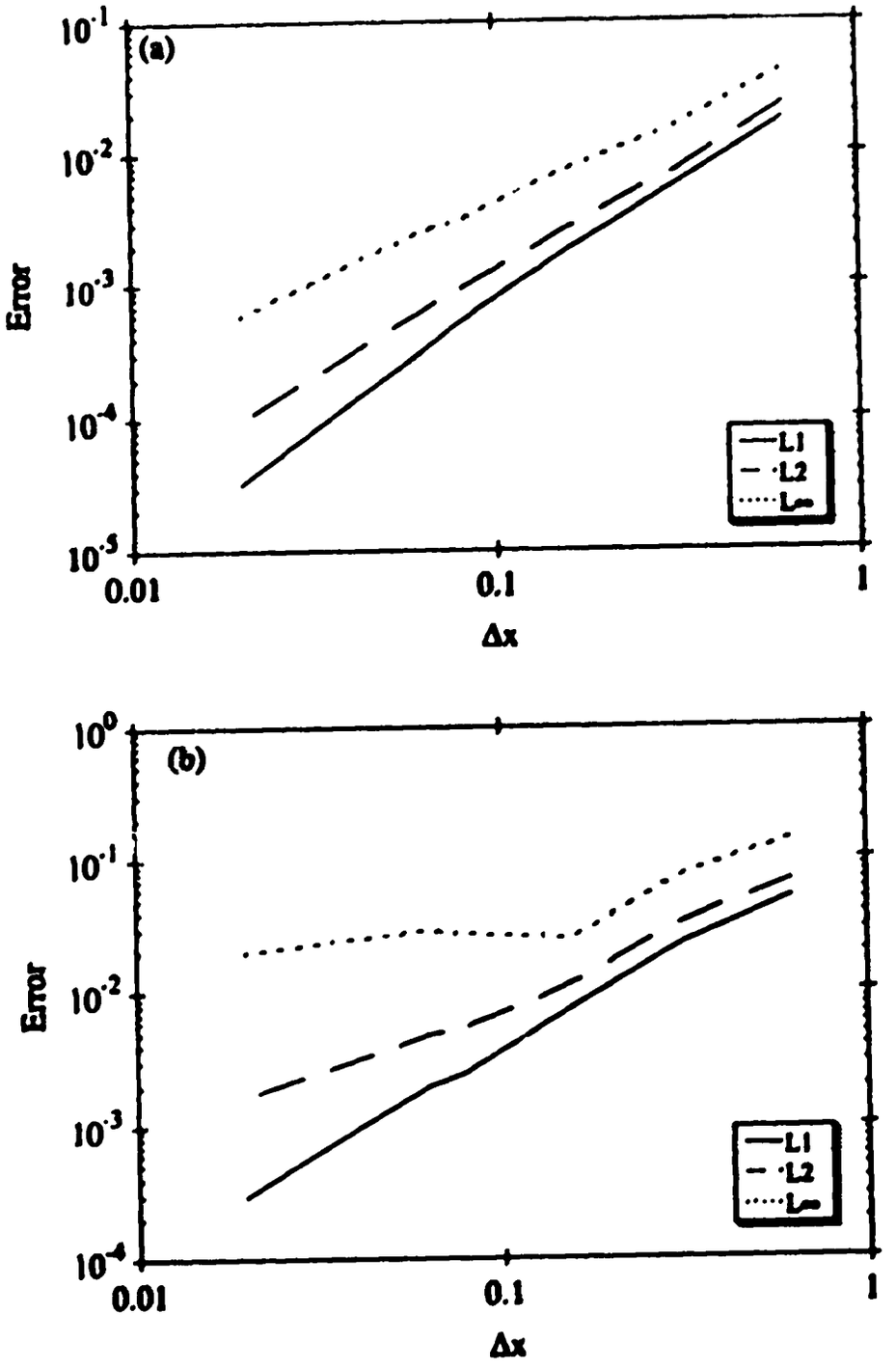


Figure 6.8: The quadratic scheme has better accuracy in general than the symmetric scheme, but after the shock forms the "knee," the solution is somewhat more severe in nature. For a small range of Δx's the solution actually diverges.

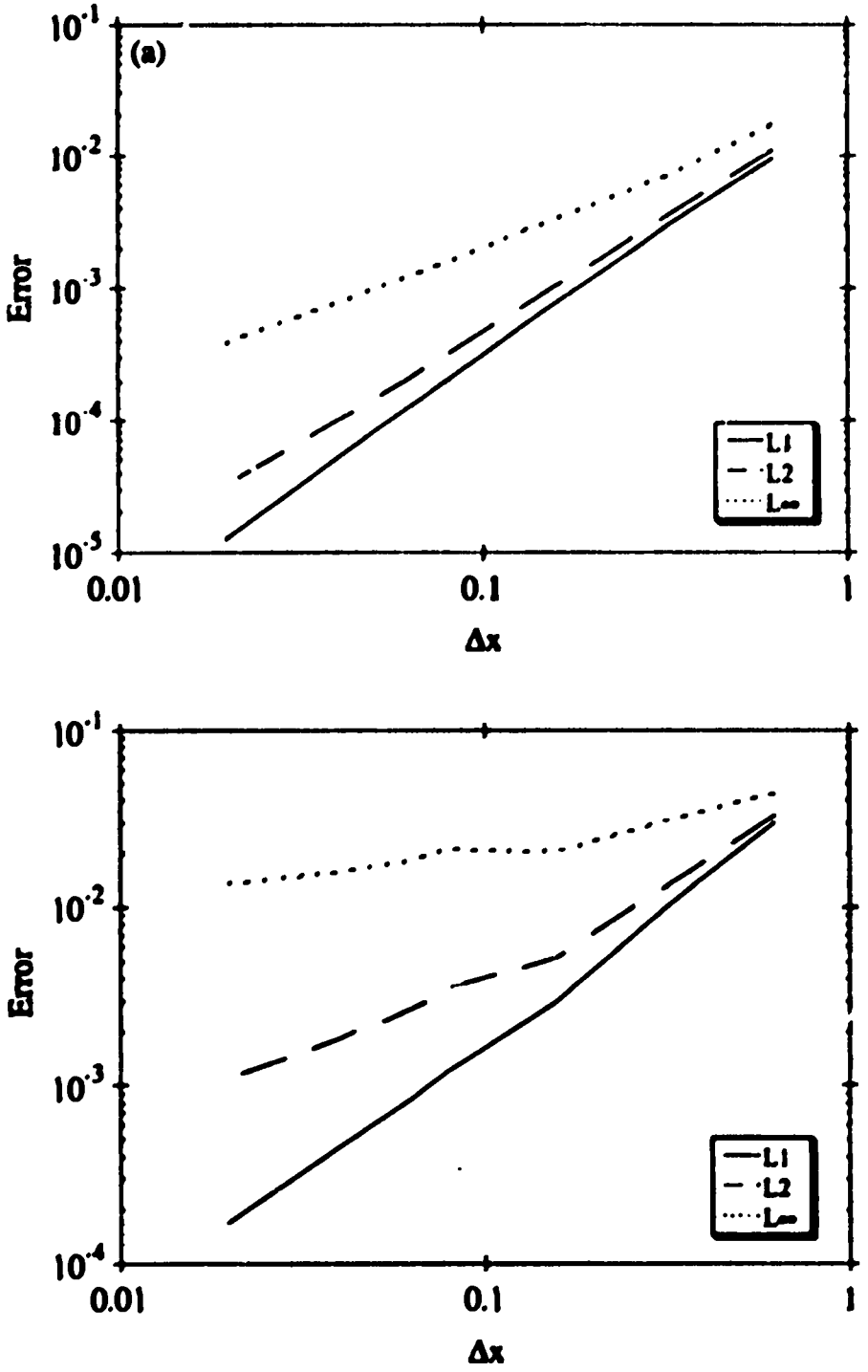


Figure 6.9: The symmetric UNO scheme has better accuracy than either of the previous methods. The convergence after the shock in the L_∞ norm is worse, however.

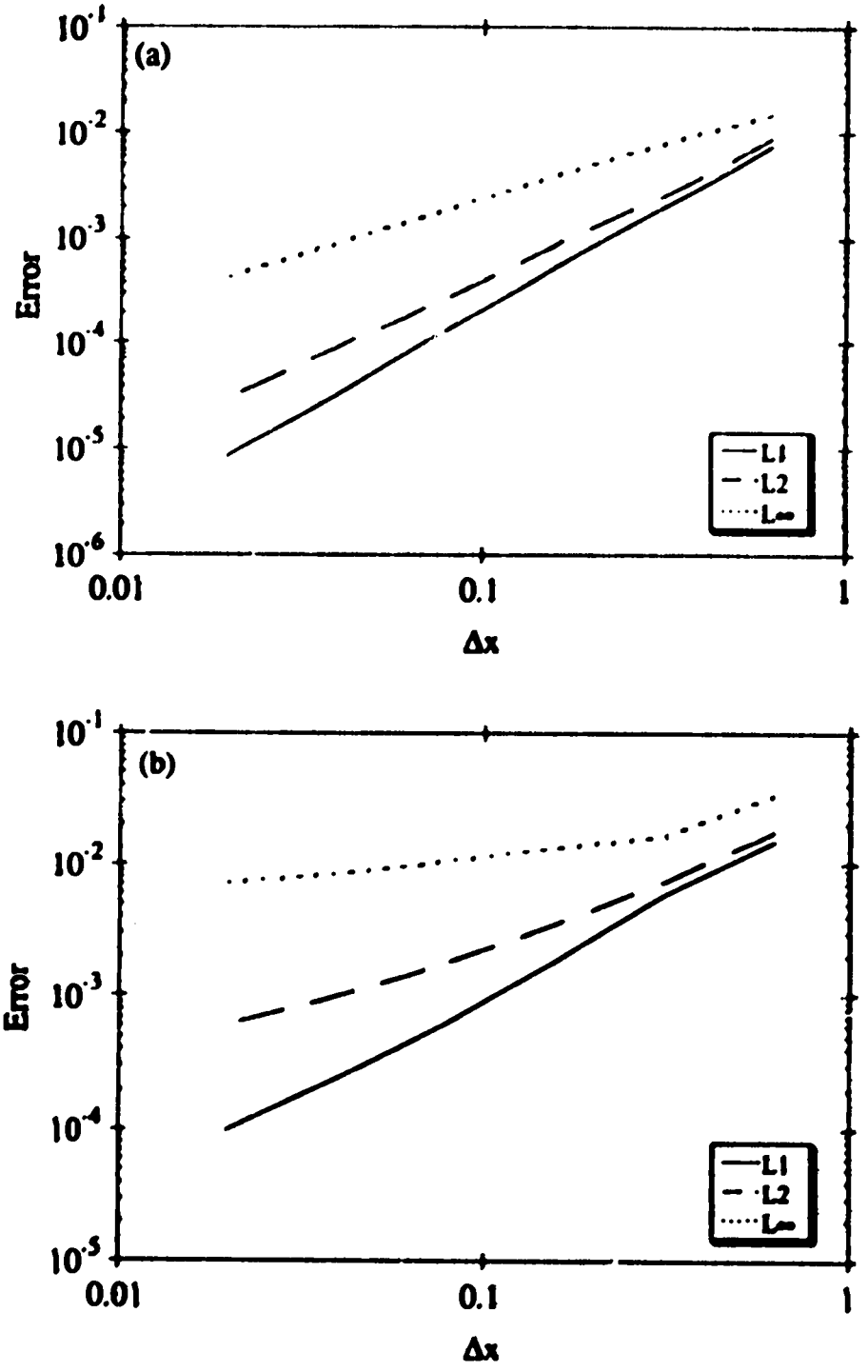


Figure 6.10: This scheme is the most accurate of the schemes shown here, but the behavior associated with the L_∞ norm at $t = 1.0$ is worse. Despite this, the solution was more accurate in every norm than any of the other methods.

Table 6.2: Order of accuracy in several norms for the schemes solving Burgers' equation when the solution contains a shock.

Scheme	L_1	L_2	L_∞
Symmetric	1.48	1.19	0.78
Quadratic	1.53	1.06	0.55
Symmetric UNO	1.50	0.99	0.39
Quadratic UNO	1.39	0.89	0.36

6.3.3 Euler Equations

Two test problems are used to test the methods on the solution of systems of equations. In both cases only the density solutions is given. For the shock tube problem, an exact solution exists and is used for comparison. In the second case, a blast wave problem, no exact solution exists, therefore a converged numerical solution is used for comparison. This solution is computed using a MUSCL scheme with a Superbee limiter on the linearly degenerate field and van Leer's limiter on the two nonlinear fields (see Chapter 8). Two thousand equidistantly spaced grid points are used with a CFL number of 0.95.

The results for these problems are given in Figs. 6.11 6.14. In general, the results of the previous section hold up for these problems. The symmetric scheme (see Fig. 6.11) gives the lowest resolution results, while the quadratic UNO scheme (see Fig. 6.14) gives the best results. The symmetric UNO scheme gives good resolution, but also suffers from some nonlinear instability resulting in oscillations. These oscillations are associated with the end of rarefaction waves as shown by Fig. 6.13. Both of the quadratic methods give better resolution of shocks and contact discontinuities than their symmetric counterparts.

In the shock tube problem, the solutions are all very similar with the resolution of the contact discontinuity being the primary difference between the methods. The quadratic UNO method also improves the smearing of the rarefaction wave. In the blast wave problem, all the methods reproduce the left of the two density peaks and all of them destroy the contact discontinuity to the left of that peak. The primary differences are in the area of resolution of the right density peak and the degree of filling in of the rarefaction between the peaks. In both cases, the quadratic UNO scheme excels by comparison.

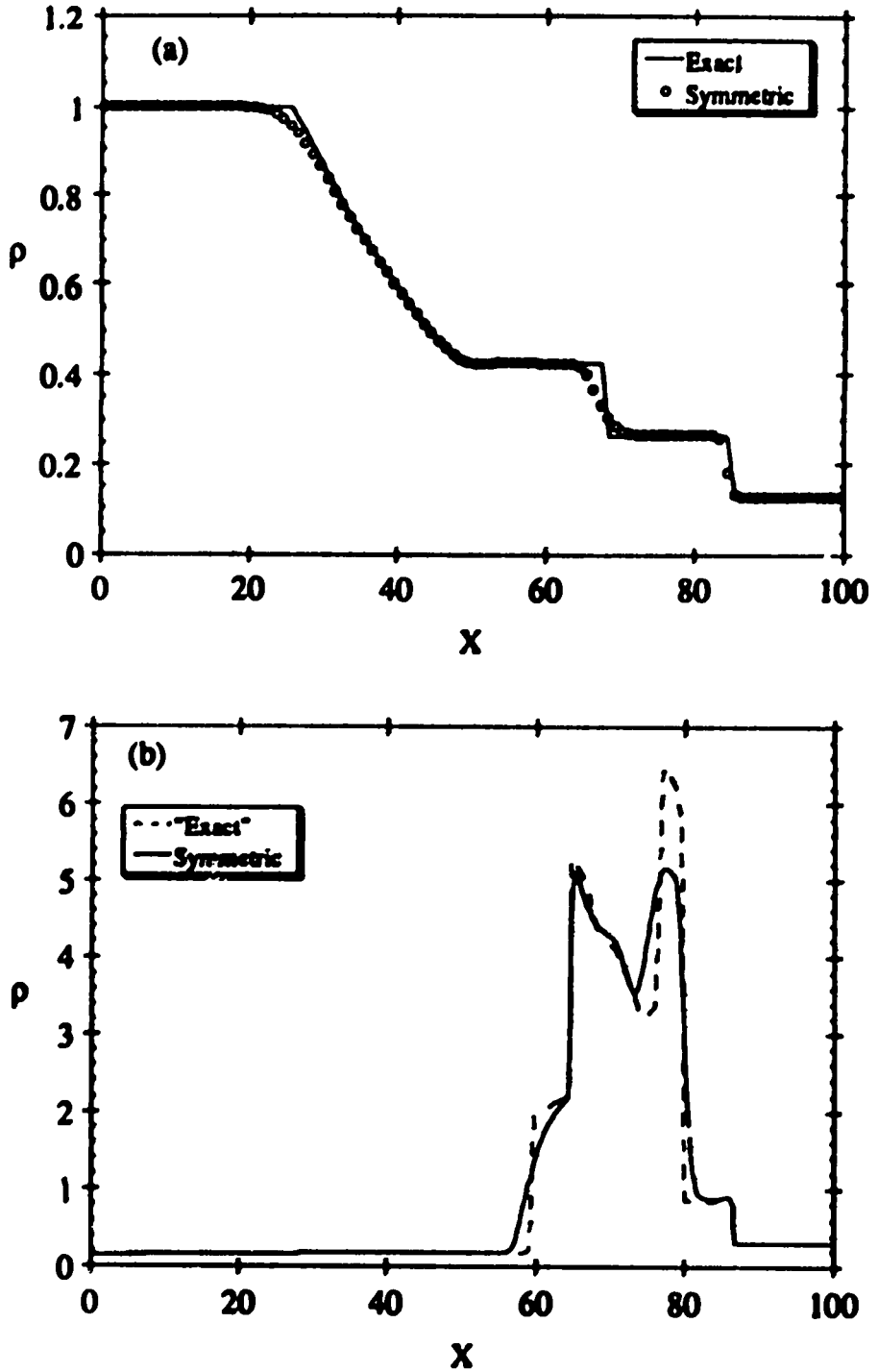


Figure 6.11: The solution of Sod's shock tube problem by the symmetric scheme is quite good except for some smearing near the contact discontinuity. The solution to the blast wave problem shows several important features also related to the smearing of contact discontinuities leading to the clipping of the right peak and the nearly complete loss of the discontinuity at $X \approx 60$. The filling in of the gap between the peaks results from smearing in rarefaction waves.

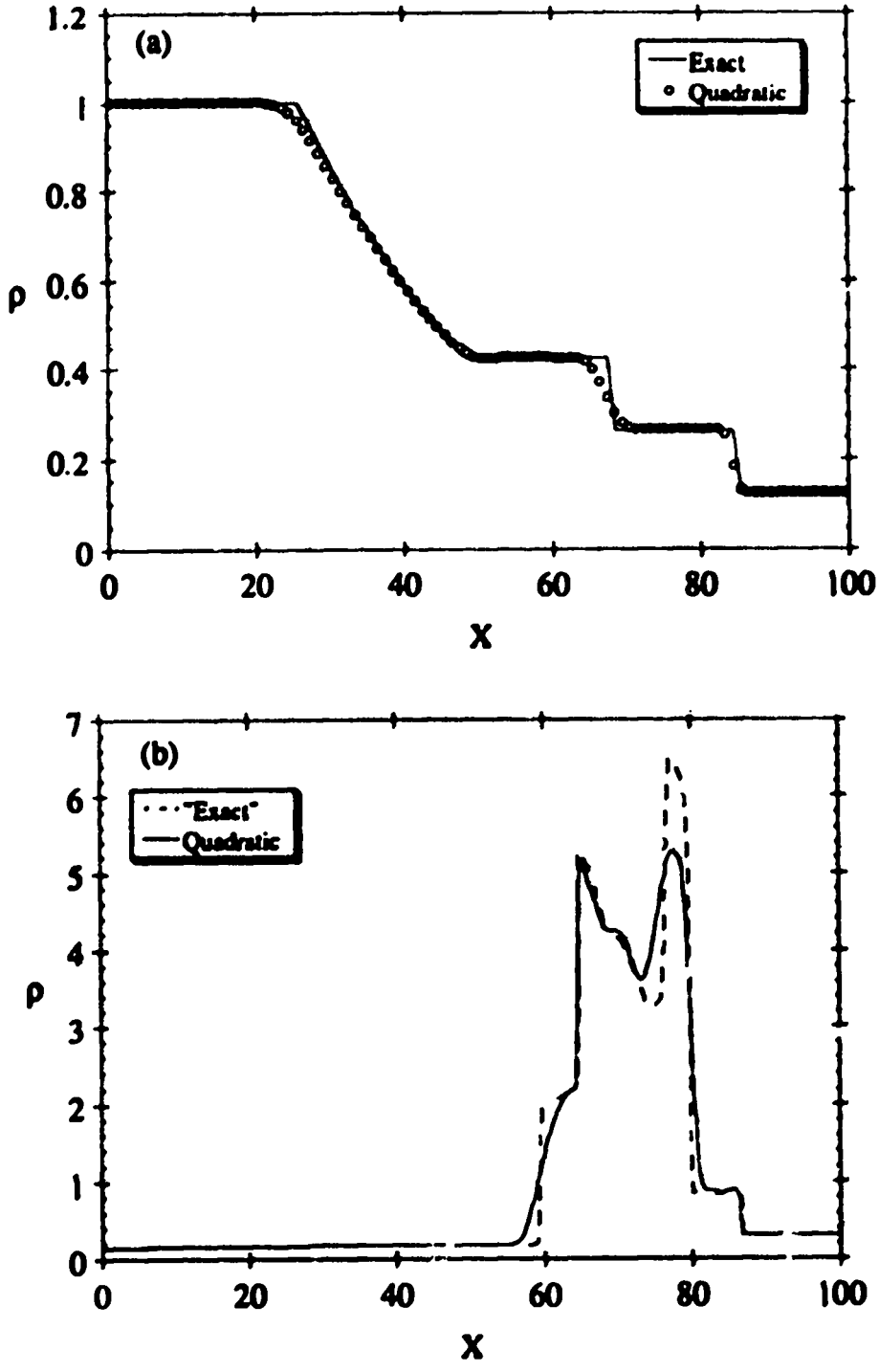


Figure 6.12: The overall results using the quadratic scheme are very similar to the symmetric scheme. The resolution of the solution is enhanced in both cases. This is especially noticeable at the shock in Sod's problem and in the left peak and rarefaction wave between the peaks in the blast wave problem.

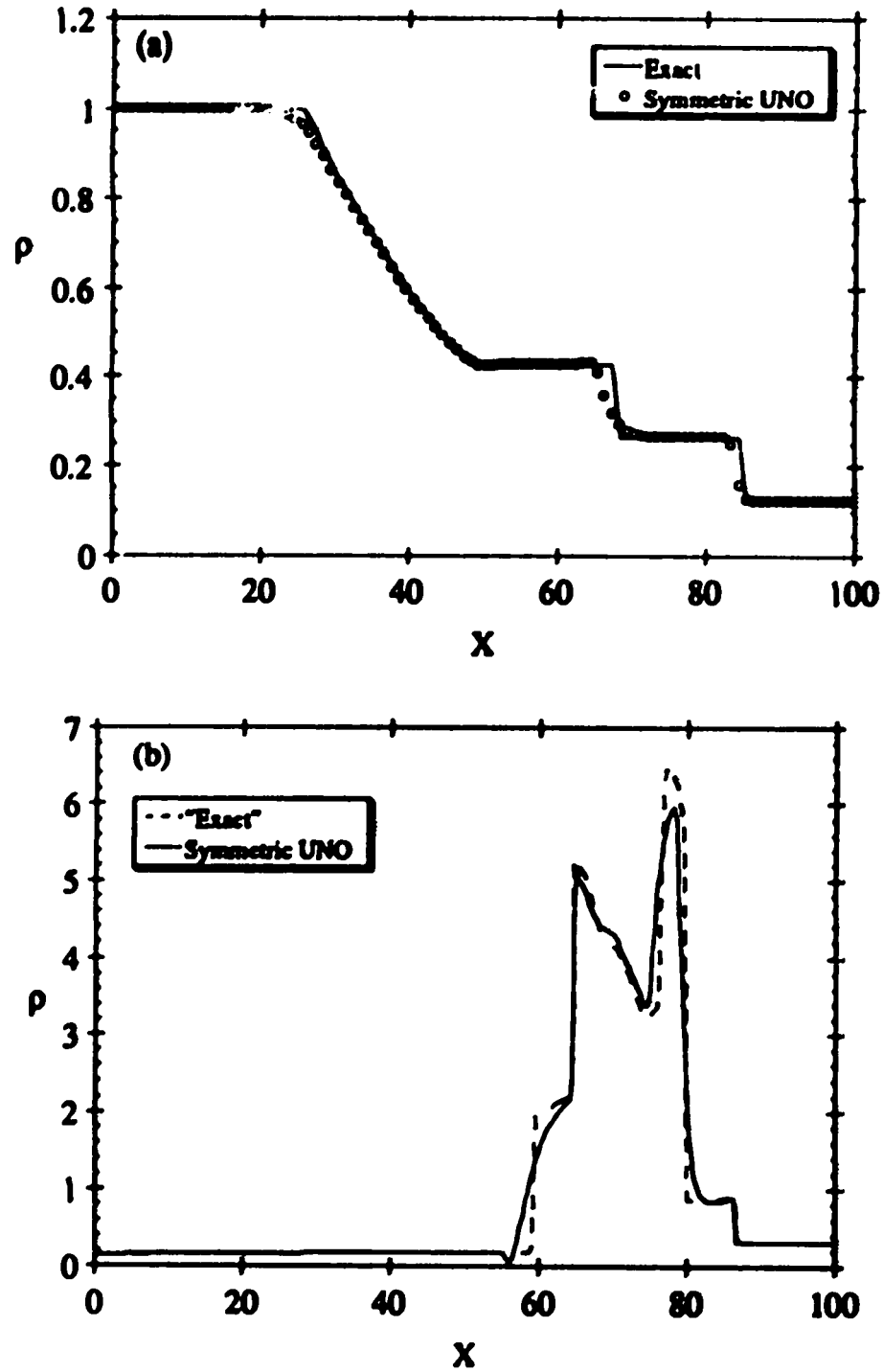


Figure 6.13: The symmetric UNO scheme gives much better resolution of contact discontinuities as shown by both figures. The price is several oscillations. One can be seen to the left of the contact discontinuity in Sod's problem. The results for the blast wave problem are quite impressive except for the dip to the left of the left-most contact discontinuity.

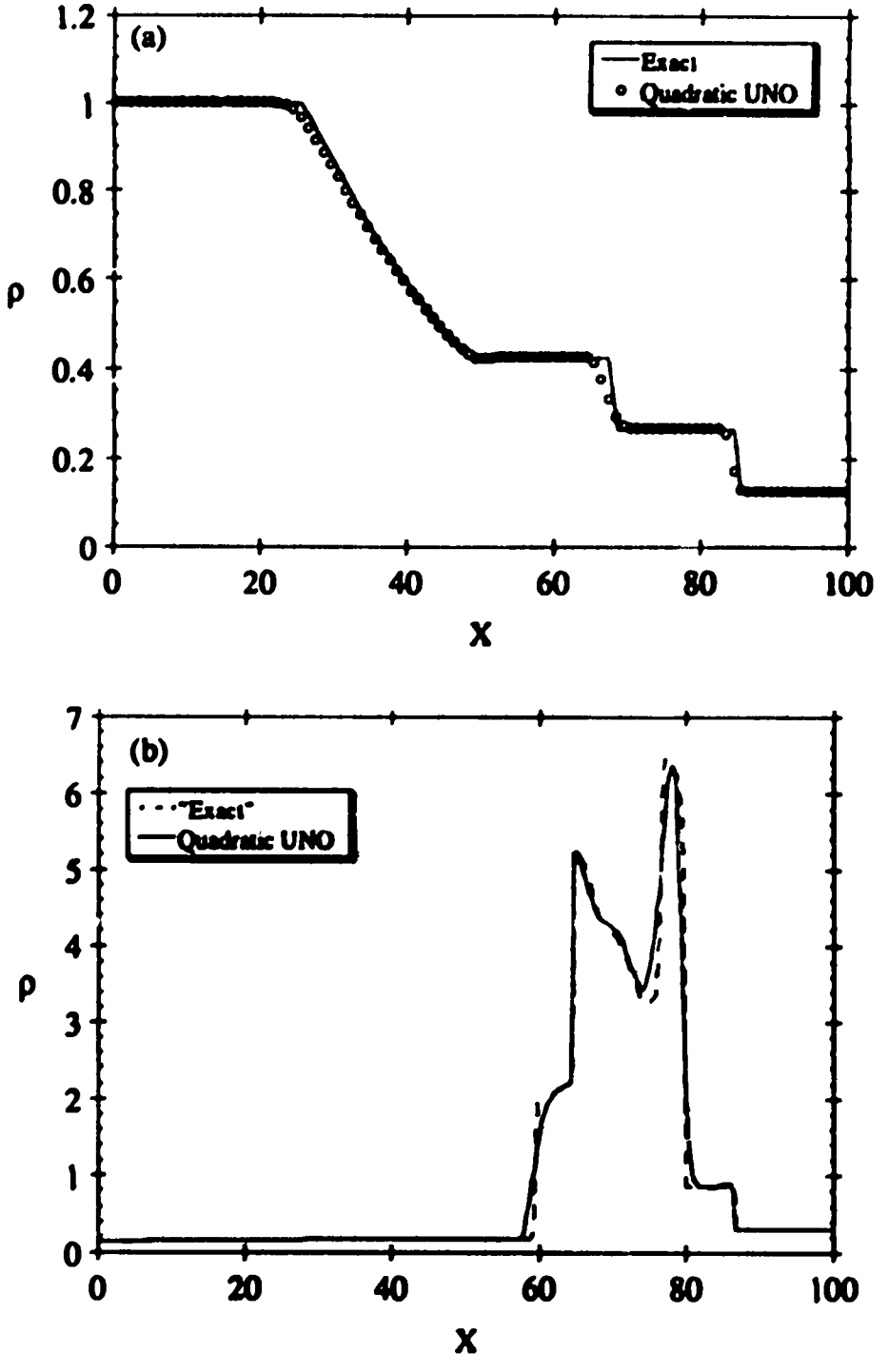


Figure 6.14: The quadratic UNO scheme seems to have the good aspects of the symmetric UNO scheme without the oscillations. For both problems, the resolution is enhanced.

6.4 Concluding Remarks

This chapter has presented an extension of the previously derived symmetric TVD methods to a geometric analog very similar to MUSCL type methods developed by van Leer. This extension has also enabled the derivation of new methods involving parabolic interpolation and the ideas of uniformly non-oscillatory methods. Through the symmetric TVD method's connection to flux corrected transport methods, these methods also tie that group of algorithms more closely to other modern algorithms.

These methods have been used to solve several test problems and have proved successful behaving as expected. Each of these newly derived method represent and improvement over the symmetric TVD method.

The topic of limiters to use with FCT methods is concentrated on in the next chapter.

Chapter 7.

FCT Limiters

A new way to pay old debts. *Phillip Massinger*

The limiters used with FCT algorithms fall into two categories: the classic type developed by Boris and Book and the generalization of Zalesak. This study started as an attempt to explain the less than stellar performance of the FCT schemes on a variety of problems and expanded in scope from there.

7.1 Classic FCT Limiters

The limiter used in the FCT methods developed by Boris and Book is nearly identical to the minmod limiter discussed in Chapter 8. The main difference is the nature of the arguments applied to the limiter. These arguments are the local gradients multiplied by the inverse grid ratio ($\Delta x/\Delta t$) and the antidiffusive flux. This makes it a three argument limiter with support identical to that found in the symmetric TVD scheme. The classic FCT limiter is

$$\mathrm{m}\left(f^a_{j+\frac{1}{2}}, \sigma^{-1}\Delta_{j-\frac{1}{2}}u, \sigma^{-1}\Delta_{j+\frac{1}{2}}u\right) . \tag{7.1}$$

This limiter can be analyzed by assuming that $f^a_{j+\frac{1}{2}} = \frac{1}{2}|a|\,\Delta_{j+\frac{1}{2}}u$ and factoring $\frac{1}{2}|a|$ out of the FCT limiter and writing the result in a ratio form

$$Q^{FCT}\left(r^-, 1, r^+\right) = \mathrm{m}\left(1, 2\nu^{-1}r^-, 2\nu^{-1}r^+\right) . \tag{7.2}$$

In this equation $r^- = \Delta_{j-\frac{1}{2}}u/\Delta_{j+\frac{1}{2}}u$ and $r^+ = \Delta_{j+\frac{3}{2}}u/\Delta_{j+\frac{1}{2}}u$. This form is equivalent to the form used for three argument TVD limiters as was discussed in Section 8.3.3. By inspection, one can see for $\nu \neq 1$ this limiter is not TVD because its result is larger than two and that the result grows infinitely large as $\nu \downarrow 0$. Figure 7.1 shows the limiter for two values of ν. The limiter is not TVD for explicit time differencing. This does not account for the stabilizing influence of the diffusive step in the solution algorithm. In Section 8.3.5, the ULTIMATE limiter is discussed. It has some similarity to the FCT limiter and as such the experience with the FCT can carry over.

As discussed in Chapter 5, this can easily be modified to rid the scheme of the need for an antidiffusive step by changing the limiter to

$$\mathrm{m}\left(f^a_{j+\frac{1}{2}}, \mu_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u, \mu_{j+\frac{1}{2}}\Delta_{j+\frac{1}{2}}u\right) . \tag{7.3a}$$

where

$$\mu = |a| \ , \tag{7.3b}$$

or

$$\mu = |a| - \nu a \ . \tag{7.3c}$$

An entropy correction as described in [182] can be applied to these definitions. This modification makes this scheme TVD and significantly improves its solutions especially for systems of equations. This formulation also allows the FCT to be used as an implicit algorithm in a similar manner as other TVD algorithms.

A second formulation based around the modified flux TVD schemes was also given in Chapter 5,

$$\begin{aligned} \text{minmod}(a,b,n) = \text{sign}(a)\max\big[\,0, &\min(n\,|a|\,, \text{sign}(a)\,b), \\ &\min(|a|\,, n\ \text{sign}(a)\,b)\big]\ , \end{aligned} \tag{7.4}$$

which for $n = 2$ gives the superbee limiter developed by Roe [176]. To get the implementation correct in the sense of a FCT method this becomes

$$\begin{aligned} \text{minmod}(n) = \text{sign}\left(f^a_{j+\frac{1}{2}}\right)\max\Big[\,0, &\min\left(\frac{1}{2}n\left|f^{AD}_{j+\frac{1}{2}}\right|, n\,\text{sign}\left(f^a_{j+\frac{1}{2}}\right)\sigma_{j-\frac{1}{2}}\Delta_{j-\frac{1}{2}}u\right), \\ &\min\left(n\sigma_{j+\frac{1}{2}}\left|\Delta_{j+\frac{1}{2}}u\right|, \frac{1}{2}n\,\text{sign}\left(f^a_{j+\frac{1}{2}}\right)f^{AD}_{j-\frac{1}{2}}\right)\Big]\ . \end{aligned} \tag{7.5}$$

This scheme is closer to the modified flux TVD formulation and produces a family of limiters shown in Fig. 5.1.

7.2 Zalesak's Generalization

Zalesak [62] redefined the FCT limiter to make it more general. The resulting limiter is nearly identical to the original FCT limiter in one dimension, but has a true multidimensional form. Zalesak also made the prescription of the antidiffusive fluxes more general, with the definition being simply stated as the difference between the low- and high-order fluxes, $f^a_{j+\frac{1}{2}} = f^H_{j+\frac{1}{2}} - f^L_{j+\frac{1}{2}}$. The low-order flux, $f^L_{j+\frac{1}{2}}$, could be any monotone numerical flux and the high-order flux, $f^H_{j+\frac{1}{2}}$, could be specified by any high-order flux.

Algorithm 3 *[Zalesak's flux limiter [62]]*

1. Sum all antidiffusive fluxes going into, A^+_j, and out of, A^-_j, a cell. In one dimension this is expressed as

$$A^+_j = \max\left(f^a_{j-\frac{1}{2}}, 0\right) - \min\left(f^a_{j+\frac{1}{2}}, 0\right)\ . \tag{7.6a}$$

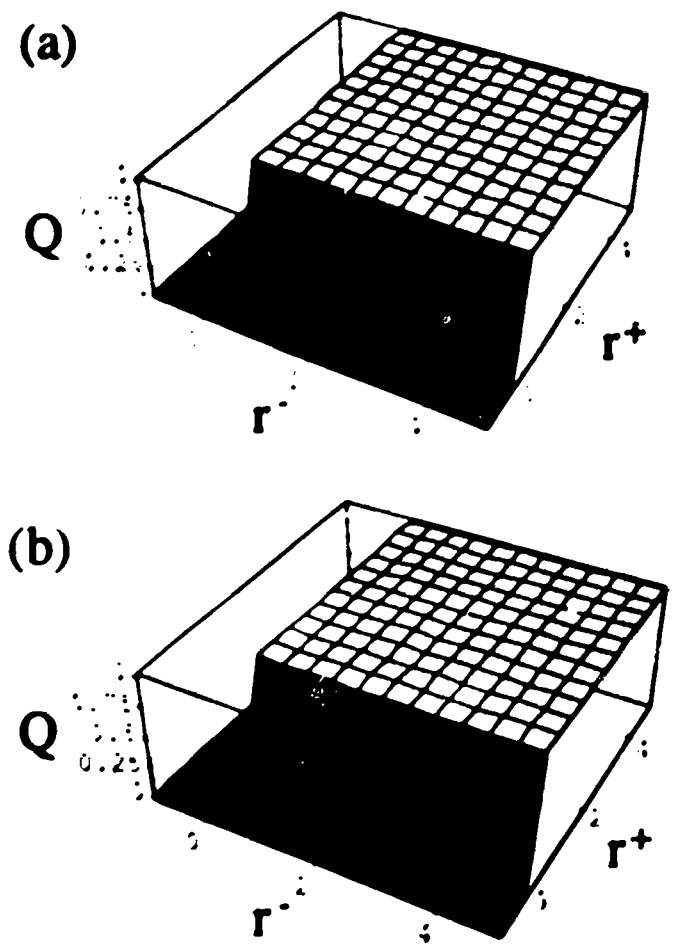


Figure 7.1: The classic FCT limiter is shown for $\nu = 0.25$ in Fig. 7.1a and $\nu = 0.5$ in Fig. 7.1b. Both of these figures show that where $r^{\pm} < 1$ the limiter is very compressive, but not second order in nature.

and

$$A_j^- = \max\left(\hat{f}^a_{j+\frac{1}{2}}, 0\right) - \min\left(\hat{f}^a_{j-\frac{1}{2}}, 0\right) . \tag{7.6b}$$

2. Find the maximum, u_j^{max}, and minimum, u_j^{min} values locally, and define

$$M_j^+ = \sigma^{-1}\left(u_j^{max} - \tilde{u}_j\right) , \tag{7.6c}$$

and

$$M_j^- = \sigma^{-1}\left(\tilde{u}_j - u_j^{min}\right) . \tag{7.6d}$$

For example u_j^{max} and u_j^{min} could be computed with the following relations:

$$u_j^{max} = \max\left(\tilde{u}_{j-1}, \tilde{u}_j, \tilde{u}_{j+1}\right) \tag{7.6e}$$

and

$$u_j^{min} = \min\left(\tilde{u}_{j-1}, \tilde{u}_j, \tilde{u}_{j+1}\right) \tag{7.6f}$$

3. Compute

$$R_j^+ = \mathrm{m}\left(1, M_j^+ / A_j^+\right) , \tag{7.6g}$$

and

$$R_j^- = \mathrm{m}\left(1, M_j^- / A_j^-\right) . \tag{7.6h}$$

4. At each cell edge, k, on the cell, j, compute

$$C_k = \min\left(R_j^+, R_i^-\right) , \tag{7.6i}$$

if $\hat{f}_k^A \geq 0$, otherwise compute

$$C_k = \min\left(R_i^+, R_j^-\right) . \tag{7.6j}$$

5. Finally, $\hat{f}_k^C = C_k \hat{f}_k^A$.

6. Zalesak also states some quality-enhancing corrections based on previous experience with the FCT

$$C_{j+\frac{1}{2}} = 0 , \tag{7.6k}$$

if

$$\hat{f}^a_{j+\frac{1}{2}}\left(\tilde{u}_{j+1} - \tilde{u}_j\right) < 0 , \tag{7.6l}$$

and

$$\hat{f}^a_{j+\frac{1}{2}}\left(\tilde{u}_j - \tilde{u}_{j-1}\right) < 0 \text{ or } \hat{f}^a_{j+\frac{1}{2}}\left(\tilde{u}_{j+2} - \tilde{u}_{j+1}\right) < 0 . \tag{7.6m}$$

The modifications made in the previous section can be applied to this limiter rather easily with by changing σ^{-1} in step 2 to σ as defined in (7.3b) or (7.3c). This

change also allows the diffusive first step to be avoided without negative consequences. The resulting algorithm is given below.

Algorithm 4 *[Zalesak's modified flux limiter]*

1. Sum all antidiffusive fluxes going into, A_j^+, and out of, A_j^-, a cell.
2. Find the maximum, $u_j^{\max}$, and minimum, $u_j^{\min}$ values locally, and define

$$M_j^+ = \mu\left(u_j^{\max} - u_j^n\right) . \tag{7.7a}$$

and

$$M_j^- = \mu\left(u_j^n - u_j^{\min}\right) . \tag{7.7b}$$

3. Compute

$$R_j^+ = \mathrm{m}\left(1., M_j^+/A_j^+\right) , \tag{7.7c}$$

and

$$R_j^- = \mathrm{m}\left(1., M_j^-/A_j^-\right) . \tag{7.7d}$$

4. At each cell-edge, k, on the cell, j, compute

$$C_k = \min\left(R_r^+, R_l^-\right) , \tag{7.7e}$$

if $\dot{f}_k^A \geq 0$ (the antidiffusive flux $\dot{f}_k^H - \dot{f}_k^L$), otherwise compute

$$C_k = \min\left(R_l^+, R_r^-\right) . \tag{7.7f}$$

5. Finally, $\dot{f}_k^C = C_k \dot{f}_k^A$.
6. Use the quality corrections substituting u_j for $\tilde{u}_j$.

Lemma 1 *For a second-order spatially accurate high-order flux, the Zalesak's modified flux limiter produces a scheme equivalent to a symmetric TVD scheme with a Q function of*

$$Q_{j+\frac{1}{2}}^{FCT} = m\left(2\mu\Delta_{j-\frac{1}{2}}u, \mu\Delta_{j+\frac{1}{2}}u, 2\mu\Delta_{j+\frac{1}{2}}u\right) . \tag{7.8}$$

Proof. For μ defined by (7.3b), the appropriate high-order flux is the second-order central difference flux. For μ defined by (7.3c) it would be the Lax-Wendroff flux. For both cases,

$$\dot{f}_{j+\frac{1}{2}}^c = \mu\frac{1}{2}\Delta_{j+\frac{1}{2}}u , \tag{7.9}$$

if the antidiffusive flux. When u_j is a local maximum or minimum, then the limiter produces a value of zero. I proceed assuming that u is monotone and increasing on the interval $[x_{j-1}, x_{j+2}]$. This interval is also used to determine u_j^{min} and u_j^{max}. The case where u is monotone decreasing is similar. Considering cell edge $j+\frac{1}{2}$, $f_{j+\frac{1}{2}}^a > 0$, thus I must find R_{j+1}^+ and R_j^-. In this case $A_j^- = f_{j+\frac{1}{2}}^a$ and $A_{j+1}^+ = f_{j+\frac{1}{2}}^a$, $(A_j^- = A_{j+1}^+)$. Because u is monotone increasing, $u_j^{min} = u_{j-1}^n$ and $u_{j+1}^{max} = u_{j+2}^n$; thus $M_{j+1}^+ = \Delta_{j+\frac{1}{2}} u$ and $M_j^- = \Delta_{j-\frac{1}{2}} u$. From these relations and the formulas for R_j^- and R_{j+1}^+, it can be seen that

$$C_{j+\frac{1}{2}} = \min\left(1, \frac{M_j^-}{A_j^-}, \frac{M_{j+1}^+}{A_{j+1}^+}\right) . \tag{7.10a}$$

Inspection shows that the terms in this limiter are identical to those asserted if the limiter is written in ratio form. When combined with the conditions for a local minimum or maximum, the minmod limiter is:

$$C_{j+\frac{1}{2}} = \frac{1}{2}\mathrm{m}\left(1, 2r^-, 2r^+\right) . \tag{7.10b}$$

By checking the form of the symmetric TVD schemes, it can be seen that this has the form of an upwind flux plus some second-order centrally differenced high-order flux multiplied by a limiter (see Section 4.5). Subtracting the low-order flux from the symmetric TVD flux gives (for $\mu = |a| - \sigma a^2$)

$$f_{j+\frac{1}{2}}^{a,TVD} = \left[\left(\left|a_{j+\frac{1}{2}}\right| - \sigma a_{j+\frac{1}{2}}^2\right) C_{j+\frac{1}{2}}\right] \Delta_{j+\frac{1}{2}} u , \tag{7.10c}$$

equating terms gives the desired result. A similar result is obtained with $\sigma = |a|$. □

Remark 22 *For higher order spatially accurate fluxes, the quality factors imposed at the end of the limiter become important (see Algorithm 3). These factors make sense in a heuristic way and definitely improve the limiters performance, but the properties of limiter are more difficult to determine in this case, although it appears to be TVD from experimental evidence. For the second-order case discussed in the previous lemma, these factors are immaterial.*

This scheme is TVD in one dimension under the conditions stated in the following theorem:

Theorem 8 *Zalesak's modified flux limiter with a second-order spatially accurate high-order flux is TVD under the following conditions*

1. *The values of u_j^{max} and u_j^{min} are taken from the set of points u_{j-1}^n, u_j^n, and u_{j+1}^n.*
2. *For σ defined by (7.3b), $|\nu| \leq \frac{1}{2}$.*
3. *For σ defined by (7.3c), $|\nu| \leq 1$.*

Proof. The conditions for a scheme to be TVD are given in Theorem 6. Using the results from Lemma 4, the proof can proceed from the standpoint of proving that a given limiter produces a TVD scheme. To ease the analysis, Zalesak's limiter is written in the form equivalent to a symmetric TVD scheme (see Lemma 4):

$$C_{j+\frac{1}{2}} \equiv Q_{j+\frac{1}{2}} = \min\left(2r^-, 1, 2r^+\right) , \tag{7.11a}$$

where $r^\pm = M^\pm/A^\pm$ with A and M defined by the modified FCT flux limiting algorithm. As given in [131], the conditions for this limiter to assure a TVD algorithm are

$$Q_{j+\frac{1}{2}} \leq 2 , \tag{7.11b}$$

$$\frac{Q_{j+\frac{1}{2}}}{r^-} < \frac{2}{\nu} - 2 , \tag{7.11c}$$

$$\frac{Q_{j+\frac{1}{2}}}{r^+} < \frac{2}{\nu} - 2 , \tag{7.11d}$$

and

$$\nu \leq 1 . \tag{7.11e}$$

These conditions should be compared with those given in Section 8.3.3. The condition (7.11e) is easily met as is (7.11b), regardless of the definition of μ. For $\mu = |a|$, the conditions of (7.11c) and (7.11d) result in a limiting CFL number of $\nu \leq \frac{1}{2}$. When $\mu = |a| - \nu a$ the right-hand sides of (7.11b)-(7.11d) are divided by $1 - \nu$. For the given limiter, the CFL condition now becomes $\nu \leq 1$. This completes the proof. □

Suitable generalizations can be made for implicit TVD schemes. These proofs do not extend to multiple dimensions, but provide some insight to the scheme's probable performance.

This method can also be applied to HOG schemes by extending the generalization made above to apply to the reconstruction step of Godunov's method. Low-order monotone fluxes are analogous to reconstructing u by piecewise constant functions equal to u_j. The antidiffusive fluxes could be made into "antidiffusive" gradients or the difference between higher order polynomial reconstructions and the low-order one. There is some ambiguity with the definition of the comparison gradients defined by $M_j^\pm$, but this can be rectified by several observations. These should be converted to gradients of similar definition, but in keeping with the FCT limiters of the past, these gradients should be multiplied by two. Previous FCT limiters had this effectively done by the limiter's construction and is an explanation for the highly compressive nature of FCT schemes. Low multiples can be chosen for this limiter to achieve greater dissipation. The remainder of the HOG algorithm can proceed conceptually without any changes.

Algorithm 5 *[Zalesak's HOG slope limiter]*

1. Define "antidiffusive" slopes, s^a, as $s^H - s^L$.
2. Sum all "antidiffusive" slopes going into, A_j^+, and out of, A_j^-, a cell.
3. Find the maximum, u_j^{max}, and minimum, u_j^{min} values locally, and define

$$M_j^+ = n\frac{u_j^{max} - u_j^n}{\Delta x^{max}}, \tag{7.12a}$$

and

$$M_j^+ = n\frac{u_j^n - u_j^{min}}{\Delta x^{min}}, \tag{7.12b}$$

where $1 \leq n \leq 2$ and with Δx^{max} and Δx^{min} being the appropriate distances from x_j to x_j^{max} and x_j^{min}, respectively.

4. Compute

$$R_j^+ = \mathrm{m}\left(1, M_j^+/A_j^+\right), \tag{7.12c}$$

and

$$R_j^- = \mathrm{m}\left(1, M_j^-/A_j^-\right). \tag{7.12d}$$

5. At each cell edge, k, on the cell, j, compute

$$C_k = \min\left(R_r^+, R_l^-\right), \tag{7.12e}$$

if $s_k^A \geq 0$, otherwise compute

$$C_k = \min\left(R_l^+, R_r^-\right). \tag{7.12f}$$

6. Finally, $s_k^C = C_k s_k^A$.

Theorem 9 *Zalesak's HOG slope limiter is TVD under the following conditions and the values of u_j^{max} and u_j^{min} are taken from the set of points u_{j-1}^n, u_j^n, and u_{j+1}^n.*

Proof. The proof is nearly identical to that given in Theorem 8, but uses the generalization of symmetric TVD schemes to a HOG formulation (see Chapter 6). □

7.3 Results

This section presents results for some of the limiters described in the pervious sections. The results are limited to the scalar wave equation and Burgers' equation. No attempt is made to present results for all the limiters given above, but the types of limiters introduced here are discussed with regard to their performance in relation to resolution and convergence. Table 7.1 shows a list of the limiters considered in the results and the abbreviations used in referring to them below.

Table 7.1: Abbreviations for the methods used in this study.

Limiter	Equation	Abbreviation
Classic FCT	(7.1)	FCTC
Zalesak's FCT	(7.6a)-(7.6m)	FCTZ
Modified FCT	(7.3a)	FCTM
Modified Zalesak's FCT	(7.7a)-(7.7f)	FCTZN

7.3.1 The Scalar Wave Equation

In this section using various limiters, the scalar wave equation is solved by the methods described in this chapter. Two initial conditions are used for the analysis: a square wave with a width of 10 cells and a $\sin^2 x$ wave (half of a period) of a width of 25 cells. Both tests are conducted for 500 time steps with a CFL number of one-half. The advective velocity is taken to be unity.

For the FCT type limiters, a Lax-Wendroff flux is used for the high-order flux in each case. In general, the FCT schemes all compete quite well with the best of the three argument limiter-based solutions. The changes required to make either the classic or Zalesak's limiter TVD result in small drop in resolution, but it is hardly noticeable. It should be stated that each FCT scheme is TVD for the cases shown. One problem that seems to plague all the three argument limiter-based schemes is the qualitative shape of the convected profile (its lack of symmetry). The FCT-based solutions seem to aggravate this problem somewhat when compared with more classic TVD solutions. Other results are given in Tables 7.2-7.4. The numerical viscosity results are explained fully in the following chapter.

A simple change to the FCT limiter can result in a large payoff. By making the limiter upwind biased, the performance of the scheme improves dramatically (this is explored in more detail in the next chapter). Staying with the scalar wave equation with $a > 0$ the classic FCT limiter would become

$$\mathrm{m}\left(f^{\circ}_{j+\frac{1}{2}}, \sigma^{-1}\Delta_{j-\frac{1}{2}}\tilde{u}\right) . \tag{7.13a}$$

and Zalesak-type limiter would only need modify the choice of C_k to

$$C_k = R_i^- , \tag{7.13b}$$

if $f_k^c > 0$ and otherwise

$$C_k = R_i^+ \tag{7.13c}$$

Table 7.2: L_1 error norms with minimum and maximum values for the square wave problem.

Limiter	Minimum	Maximum	L_1 error
FCTC	0.0000	0.8376	5.85×10^{-2}
FCTZ	0.0000	0.8310	5.95×10^{-2}
FCTM	0.0000	0.7923	6.35×10^{-2}
FCTZN	0.0000	0.7782	6.42×10^{-2}
FCTCU	0.0000	0.8377	5.85×10^{-2}
FCTZU	-0.0522	0.8899	5.75×10^{-2}
FCTMU	0.0000	0.8096	5.99×10^{-2}
FCTZNU	0.0000	.8090	5.99×10^{-2}

Table 7.3: L_1 error norms with minimum and maximum values for the $\sin^2 x$ wave problem.

Limiter	Minimum	Maximum	L_1 error
FCTC	0.0000	0.9509	2.91×10^{-2}
FCTZ	0.0000	0.9511	2.99×10^{-2}
FCTM	0.0000	0.9556	2.93×10^{-2}
FCTZN	0.0000	0.9523	3.00×10^{-2}
FCTCU	0.0000	0.9514	2.92×10^{-2}
FCTZU	-0.0278	0.9716	3.22×10^{-2}
FCTMU	0.0000	0.9587	3.02×10^{-2}
FCTZNU	0.0000	0.9587	3.02×10^{-2}

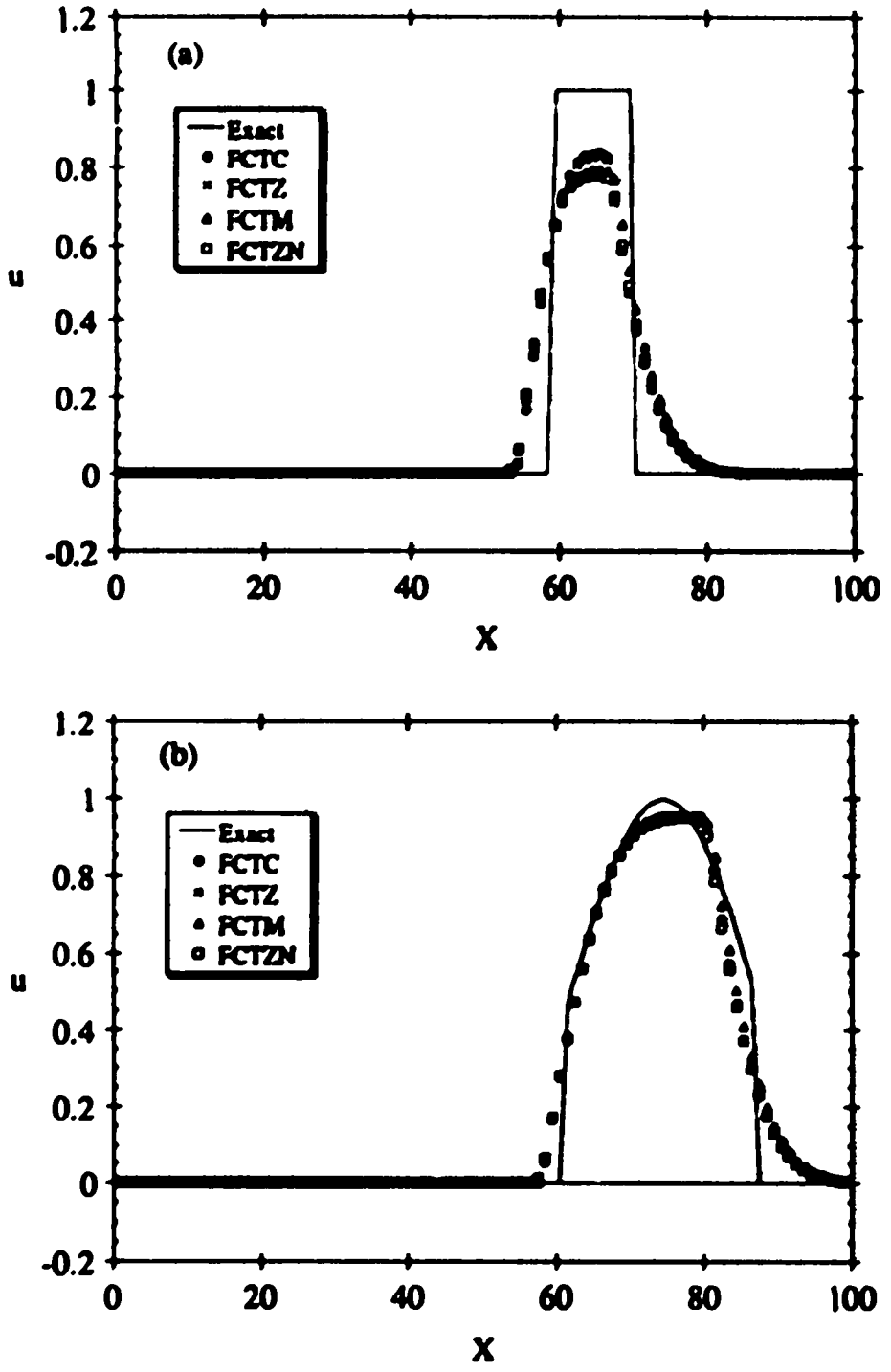


Figure 7.2: The scalar square and $\sin^2 x$ wave solutions using several FCT limiters with a Lax-Wendroff high-order flux.

Table 7.4: Numerical viscosity and total variation for both scalar wave equation problems.

Limiter	$\sum r$ square	TV square	$\sum r \sin^2 x$	TV $\sin^2 x$
FCTC	26.67	1.68	16.99	1.90
FCTZ	27.44	1.66	17.81	1.90
FCTM	31.09	1.58	18.52	1.91
FCTZN	31.04	1.56	18.31	1.90
FCTCU	26.64	1.68	16.97	1.90
FCTZU	27.20	1.89	19.14	2.01
FCTMU	29.60	1.62	18.16	1.92
FCTZNU	29.60	1.62	18.16	1.92

These schemes are denoted by the same nomenclature as used above, but with a "U" at the end of the acronym. For the classic FCT limiter the effect of this change is minimal. For Zalesak's limiter, the impact makes the solution oscillatory. For the modified limiters there is an improvement for the square wave problem, but the $\sin^2$ problem the effects wash out. The tabular data reflects this, as does Fig. 7.3.

7.3.2 Burgers' Equation

This section of the chapter centers around the order of accuracy obtained with methods in conjunction with limiters and their subsequent solutions. To accomplish this, a standard test problem using Burgers' equation is used. The problem consists of an initial condition of $\sin(x)$, $x \in [0, 2\pi]$. At $t = 0.2$, the solution is smooth, and at $t = 1.0$, a shock has formed in the solution. It is at these times that the accuracy of the solution is assessed. The problem is solved with 10 grid cells followed by 1000 grid cells.

The results for this test problem are given in Tables 7.5 and 7.6. The FCT limiters seem to suffer from poor convergence characteristics. In general, the modified FCT limiters are more efficient and provide resolution on course grids.

7.4 Concluding Remarks

In this chapter a number of limiters have been reviewed and their properties examined. In addition, several limiters have been introduced or reformulated and analyzed within a common framework. The impact of limiters on high-resolution numerical

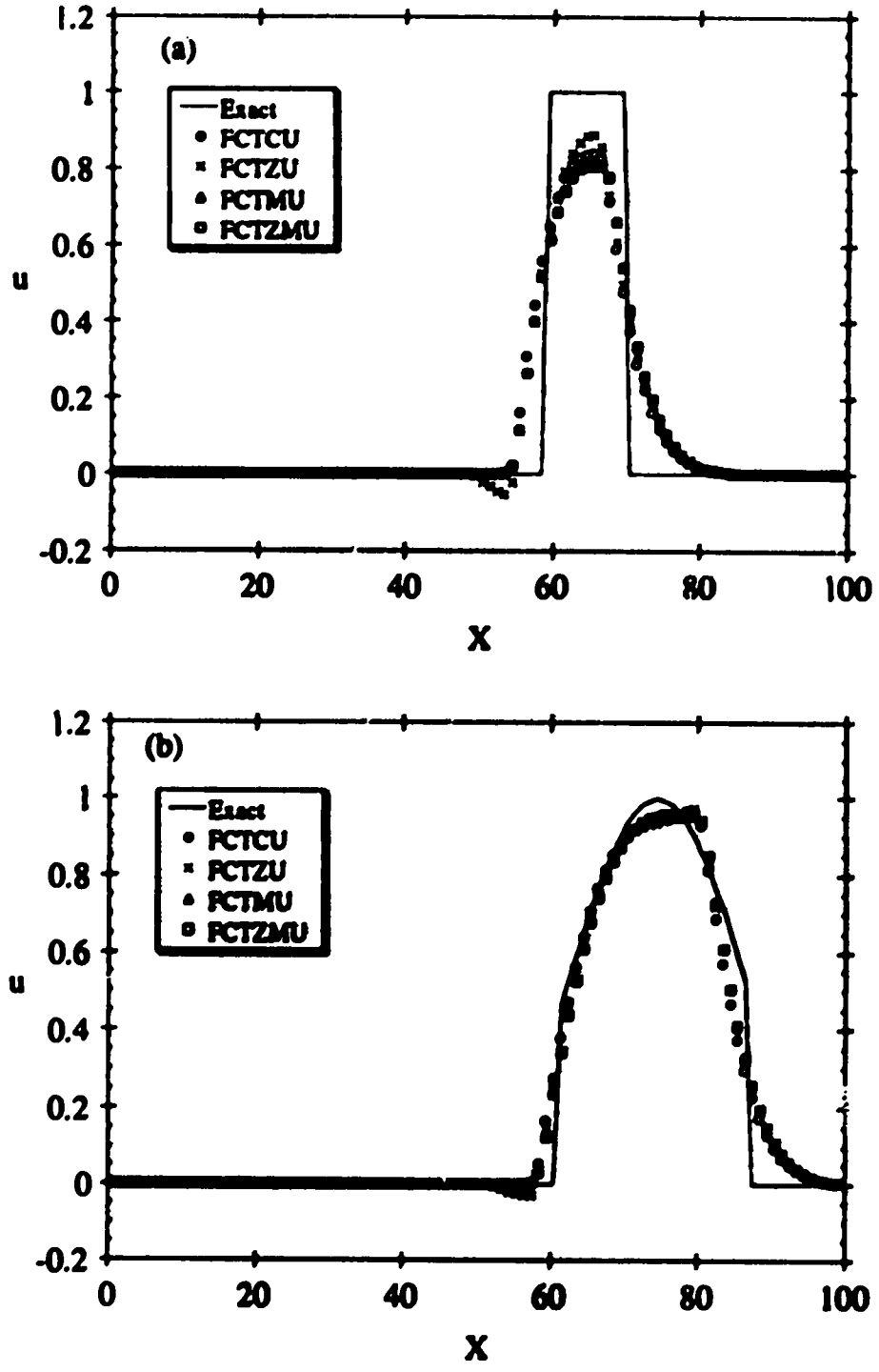


Figure 7.3: The scalar square and $\sin^2 x$ wave solutions using several FCT limiters with a Lax-Wendroff high-order flux and upwind biasing.

Table 7.5: Order of convergence in several error norms for Burgers' equation at $t = 0.2$ when the solution is smooth.

Limiter	L_1	L_2	L_∞
FCTC	2.00	2.01	1.74
FCTZ	1.97	1.67	1.13
FCTM	1.87	1.58	1.12
FCTZN	1.91	1.58	1.08

Table 7.6: Order of convergence in several error norms for Burgers' equation at $t = 0.2$ when the solution has a shock in it.

Limiter	L_1	L_2	L_∞
FCTC	1.42	0.89	0.33
FCTZ	1.46	0.91	0.33
FCTM	1.49	0.94	0.37
FCTZN	1.34	0.80	0.28

solutions has also been demonstrated. The importance of limiters on the solution of the equations is undeniable. The quality of solutions is directly traceable to the limiters because they are the heart of the numerical schemes.

More study of limiters is warranted in light of these results. As discussed earlier, limiters can impact steady-state solution convergence. Some study of this phenomena is needed. Additionally, both TVB and generalized average limiters should studied in order to give more systematic manner to choose the constants used with the limiters.

The next chapter explores the topic of limiter more generally and in more detail.

Chapter 8.

TVD and Nearly TVD Limiters

The road to resolution lies by doubt. *Francis Quarles*

8.1 Background

Godunov gave the impetus for the development of modern high-resolution methods with his paper [56]. Boris and Book [59] realized that Godunov's theorem meant that a second-order "monotone" algorithm could be constructed if it were nonlinear in nature. In deriving their FCT algorithm, they introduced limiters as a means to assuring second-order accuracy with "monotone" results.

8.2 Introduction

This line of thought was also followed by other pioneers in the field. Van Leer used a nonlinear limiters function in defining what has become known as the classic MUSCL algorithm [119]. Harten and Zwas used a similar formalism in deriving the hybrid method [146], as did Harten with artificial compression method [183]. The methods developed by van Leer and Harten took the form of switching functions between high- and low-order schemes. Thus the high-order scheme would be used where the solution is smooth, and the low-order solution is used near discontinuities to guard against the formation of oscillations.

Van Leer extended this line of thought more directly to a high-order extension of Godunov's method in [120, 60]. The limiters were used to define polynomial reconstructions of the dependent variables used to derive difference approximations for the numerical fluxes. This general line of thought led to schemes known as HOG schemes. These schemes can be viewed similarly to the switching schemes discussed previously. The limiters are used to blend high- and low-order approximations guarding against oscillations. The major difference is the inclusion of the Riemann problem in the solution scheme, thus embodying the essence of upwind weighted differencing.

The general form of limiters defined in the FCT schemes and by van Leer's HOG schemes were used to define TVD schemes. Harten [130, 61] introduced the concept of nonlinear TVD finite difference schemes. This concept was also used by Roe [131, 176], Sweby [132], and Davis [133] to define a class of schemes based on TVD corrections to the Lax-Wendroff [58] scheme. This work was summarized by Yee [134] where one member of this class of schemes was dubbed as the "symmetric TVD" scheme. In recent years, several authors have made firmer connections between FCT and

TVD/HOG methods [184, 185]. I have written about this relation in Chapters 6 and 5. In those chapters, the relation between the FCT method as stated by Zalesak and the symmetric TVD schemes and subsequently the relation to the symmetric TVD scheme to HOG type methods are explored. This line of approach can benefit all forms of high-resolution solution of hyperbolic conservation laws by adding a larger degree of synergism between these various formulations.

This chapter has been organized into four sections. The next section describes a wide variety of limiters used in the construction of high-resolution algorithms. This exposition includes material applicable to TVD and TVB schemes as well as generalizations to limiters generally denoted by the label, "nearly TVD." A number of limiters discussed in the third sections are used to solve the scalar wave equation and Burgers' equation. These results are given and discussed in the fourth section. The final section discusses conclusions.

8.3 Description of Limiters

In my opinion, this subject has been given inadequate coverage in the literature despite its relative importance to the derivation of nonoscillatory high-resolution difference schemes. Sweby [132, 186, 187] has given the most widely referenced coverage of the subject. Roe [131, 176] also gave attention to the subject. A more detailed discussion of these references is given in the following sections.

The work contained in [132] and [176] is limited to an upwind-biased limiter applied to a TVD Lax-Wendroff scheme [133, 5, 134]. Roe's work given in [131] applies to a TVD Lax-Wendroff scheme where the limiter is not biased with the wind, which has become known as the symmetric TVD schemes. Because the limiter is cell-edge centered this requires the limiter to use three arguments rather than two as in the upwind-biased case (also see [8, 6, 134]). This is significant in algorithmic performance as noted later in this chapter. Munz [181] surveyed a number of limiters with relation to a HOG scheme for a scalar two-dimensional equation using operator splitting (see Appendix F). In this work problems with both symmetry and resolution were noted with symmetric TVD schemes.

8.3.1 General Requirements

To begin the discussion of limiters, a concise definition is presented.

Definition 5 (Limiters) *A limiter is a mechanism that imposes specified constraints on the computation of the numerical flux producing higher order accuracy, but also controlling oscillations and sometimes improving the resolution of discontinuities adaptively.*

This definition fails to encompass the full range of limiters given in the literature. It does give the general concept embodied by limiters. The constraints in many cases are

taken to be the restriction to TVD discretizations of a scalar hyperbolic conservation law. Often, as is the case with the FCT, the limiter is defined in a more somewhat heuristic manner, namely to keep new extrema from being formed in the solution.

At this point, it is useful to delineate the difference between slope and flux limiters more closely. This is done from the standpoint of a philosophical differentiation rather than from a purely substantive basis. The slope limiters can be thought as being used directly during interpolation. Flux limiting usually involves methods that are classified as finite-difference types. Thus slope limiting applies to HOG algorithms and the flux limiting applies to TVD and FCT algorithms. One caveat can be placed on this classification, it is not stringent. An example of this are the ENO schemes due to Shu and Osher [65, 66, 188].

Remark 23 *In general slope limiting refers to the reconstruction (projection) phase of the solution process. Flux limiting infringes on the solution in the small (evolution) portion of the solution. In [147], van Leer admonishes this practice. The evolution process can aid in the limiting process through the determination of the domain of dependence for the limiter. This principle has been used successfully with upwind-biased cell-edge type TVD Lax-Wendroff schemes or, for that matter, linear schemes such as the Beam-Warming scheme.*

Typically, a limiter is used to choose the smoother of several gradients with some caveats imposed to improve the quality. This can also be viewed as a form of averaging which is nonlinear rather than linear in nature. The averaging can also have the condition of setting its value to zero if the arguments differ in sign. This condition with appropriate limits on the magnitude of the resultant gradient in relation to other local gradients results in "monotone" solutions. Other limits of the resultant scheme can be applied to give something closer to an ENO type of philosophy.

The limiter functions have a general form given by the "minmod" type

$$Q = \mathrm{m}(a, b) \ , \tag{8.1a}$$

or

$$Q = \mathrm{m}(a, b, c) \ , \tag{8.1b}$$

where

$$\mathrm{m}(a, b) = \mathrm{sgn}(a) \max\left[0, \min\left(|a|, \mathrm{sgn}(a)\, b\right)\right] \ , \tag{8.1c}$$

or

$$\mathrm{m}(a, b, c) = \mathrm{sgn}(a) \max\left[0, \min\left(|a|, \mathrm{sgn}(a)\, b, \mathrm{sgn}(a)\, c\right)\right] \ . \tag{8.1d}$$

This definition can easily be extended to an arbitrary number of arguments. As one can see, the minmod limiter returns the minimum of the arguments unless they differ in sign. If they differ in sign, the result is zero. As I show in Section 7.1, this form was introduced with the FCT method of Boris and Book [59].

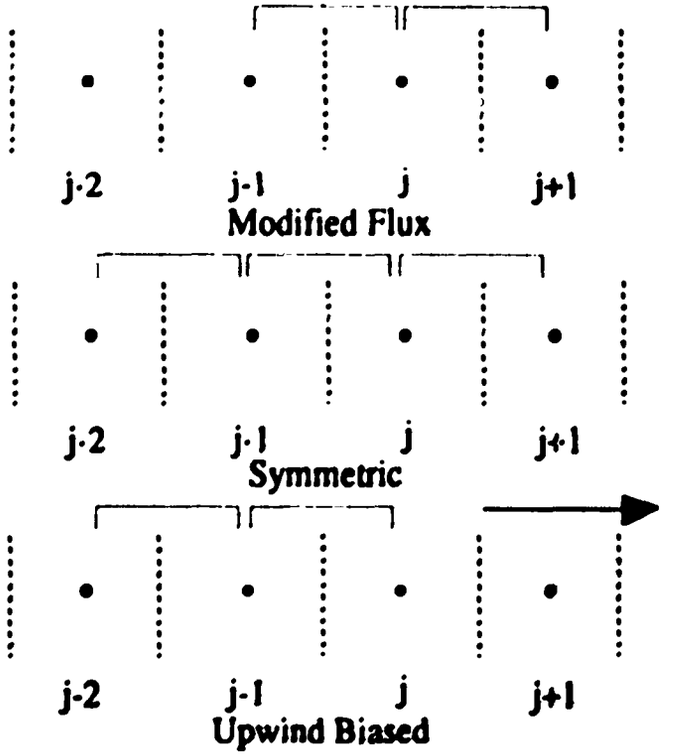


Figure 8.1: The computational stencil of the main limiter types in one dimension. Brackets indicate which points are used in evaluating local gradients. The modified flux or cell-centered limiter is centered about grid point j, the symmetric limiter is centered about cell-edge $j - \frac{1}{2}$, and the upwind-biased limiter for cell-edge $j - \frac{1}{2}$ is centered about cell $j - 1$ for $a > 0$. For $a < 0$ it would have the same stencil as the cell-centered limiter.

Limiters are centered in some sense. They can be centered about a grid point, cell edge, or biased by the direction of the flow as shown by Fig. 8.1. The appropriate definition of this centering is determined by the requirements of the underlying polynomial reconstruction. The limiters are defined at the points where a gradient of some sort is needed in the scheme definition.

Roe [176] and Sweby [132] introduced a formulation of these limiters that is particularly useful for analysis. Yee [134] also used this form in her analysis of symmetric TVD schemes. In this form, the function $Q_{j+\frac{1}{2}}$ is rewritten in terms of ratios of local gradients denoted by $r = \Delta_k u / \Delta_{j+\frac{1}{2}} u$ under this formulation. The minmod limiter has a slightly modified form

$$m(a, b) = \max[0, \min(1, r)]\, a\,, \tag{8.2}$$

with $r = b/a$, which has an similar functional form for three arguments.

Roe and Sweby also gave some desirable properties for limiters to have such as symmetry (applicable to two argument limiters)

$$\frac{Q(r)}{r} = Q\left(\frac{1}{r}\right) \ , \quad \text{or} \quad Q(a,b) = Q(b,a) \ , \tag{8.3a}$$

and homogeneity

$$Q(\mu, \mu r) = \mu Q(1, r) \ . \tag{8.3b}$$

Although the homogeneity property can easily be generalized, the symmetry property is in need of proper generalization for limiters using more than two arguments.

Another property discussed by Roe [176] is that of linear averaging. Quadratic data could be exactly advected with the use of a function of the form

$$Q(a,b) = \mu a + (1-\mu) b \ , \ \mu \in [0,1] \ , \tag{8.3c}$$

because in quadratic data the differences in gradients vary linearly. This characteristic cannot be used with TVD limiters because this would produce a linear algorithm and produce oscillatory solutions by virtue of Theorem 3. Some of the characteristics of this property can be recovered when the flow field is smooth and resolved.

Although this is not commonly stated, the limiters used in TVD schemes are convex and consistent averages of their local data's gradients. This is equivalent to stating that the schemes are second-order accurate because the limited gradients and the resulting schemes are convex averages of a family of second-order linear schemes. Thus a general form of limiters is

$$Q(a_1, a_2, \ldots, a_n) = c_1 a_1 + c_2 a_2 + \ldots + c_n a_n \ , \tag{8.4a}$$

where

$$c_j \geq 0 \ , \ j \in [1,n] \ , \tag{8.4b}$$

and

$$\sum_{j=1}^{n} c_j = 1 \ . \tag{8.4c}$$

Consistency would dictate that

$$Q(a, a, \ldots, a) = a \ . \tag{8.4d}$$

As discussed in more detail below (Sections 8.3.3 and 8.3.3), the commonly used TVD limiters have this property whereas some other limiters of similar design (such as the FCT or ULTIMATE limiters) do not.

One key point in this entire discussion is that the limiters in conjunction with upwind principles attempt to balance resolution with the need for dissipation in the algorithms. It is this trade off that is vital to the success of schemes. It is explored

in the next section.

8.3.2 Numerical Dissipation

The view can be taken that the limiter is simply a "fancy" form of artificial dissipation. This is true to a certain extent when considering the classical depiction of artificial dissipation, but the difference is that the choice of dissipation coefficients is nonlinear. To see this, I recall the observation given in [30] that an upwind-differenced scheme solves the following parabolic equation to second-order accuracy:

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = \frac{1}{2}|a|\,\Delta x\,(1-\nu)\,\frac{\partial^2 u}{\partial x^2}\,. \tag{8.5}$$

This equation can be derived by taking the difference between the numerical schemes for upwind differencing and Lax-Wendroff's method. Taking this approach a sort of numerical viscous stress can be defined as

$$\tau^{up}_{LR,N} = f^{LW}_{LR} - f^{up}_{LR}\,. \tag{8.6}$$

Using the approach outlined above for HOG-type algorithms yield a useful measure of a limiter's effect on the solution. These relations are given for a scheme defined by the following polynomial:

$$P_j(x) = u_j + \widetilde{\Delta_j u}\frac{(x-x_j)}{\Delta_j x}\,, x \in \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right] \tag{8.7}$$

where $\widetilde{\Delta_j u} = Q_j \Delta_{j+\frac{1}{2}} u$. Using a Lax-Wendroff-type time discretization and constant mesh spacing gives for upwind differencing

$$\tau_{LR,N} = \frac{\Delta x}{4}\left[(a-|a|)(1+\nu)(Q_{j+1}-1) + (a+|a|)(1-\nu)(1-Q_j)\right]\Delta_{j+\frac{1}{2}} u\,, \tag{8.8}$$

where Q is defined as

$$Q_j = \frac{\widetilde{\Delta_j u}}{\Delta_{j+\frac{1}{2}} u}\,. \tag{8.9}$$

For Lax-Friedrichs' differencing used as the underlying E-scheme gives

$$\tau_{LR,N} = \frac{\Delta x}{4}\left[\left(a - \frac{|a|}{\nu}\right)(1+\nu)(Q_{j+1}-1) + \left(a + \frac{|a|}{\nu}\right)(1-\nu)(1-Q_j)\right]\Delta_{j+\frac{1}{2}} u\,. \tag{8.10}$$

Remark 24 *For general use in computing the quantity $\tau_{LR,N}$ the difference between the Lax-Wendroff flux and a certain high-order flux is used.*

Several observations can be made by carefully analysing these functions. For an

upwind-based scheme, the viscous stress is with the gradient $\Delta_{j+\frac{1}{2}}u$ whenever the limiter gradient is taken to be the minimum gradient or less; however, if the limited gradient is larger than one of the local gradients, then the stress can be against the gradient or anti-diffusive. The second of these two cases leads to compression in an algorithm. Geometrically, the orientation of the cell averages becomes inverted at the computed cell-edge values. If this persists for many time steps, it would lead to a disastrous instability, but the nonlinear nature of the limiters guards against this occurrence.

This is of some consequence with the Lax-Friedrichs-based scheme (or similarly based schemes such as a local Lax-Friedrichs [65, 66] or the HLLE solver [130, 128]). In most cases, the diffusive effect is enhanced by the increased diffusion, but where the limiter produces an antidiffusive flux, the antidiffusive nature is enhanced by the diffusion. This can lead to small oscillations. This behavior is exemplified by the FCT limiters where the limiter has an antidiffusive Lax-Friedrichs-type signal speed (σ^{-1}).

8.3.3 TVD Limiters

Although this is not completely general, for the purposes of this study the limiters used with TVD schemes can be divided into two categories: two argument and three argument types. These limiters can also be used with FCT schemes as I have reformulated them and with HOG algorithms corresponding to a given TVD scheme. The principal contributions found in the following sections are generalizations of the ideas of Sweby [132] and Roe [176] to more general numerical schemes. The analyses of Sweby and Roe used with an upwind-biased TVD Lax-Wendroff scheme applies very well to other uses of two argument limiters. The analysis of Roe [131] with regards to three argument limiters is limited to a small set of the limiters which are a natural outgrowth of the two argument limiters.

For general second-order TVD schemes, several condition must be met for the limiters to provide a TVD solution. These are taken from the conditions for a TVD scheme in a semi-discrete case, (see Chapter 4). For cell-centered based limited schemes such as the modified flux TVD scheme in (4.22a), the conditions are for $a \geq 0$

$$\frac{Q_j}{r} - Q_{j+1} \leq 2\,, \tag{8.11a}$$

and for $a < 0$

$$Q_{j+1} - \frac{Q_j}{r} \leq 2\,. \tag{8.11b}$$

For cell-edge based limited schemes such as (6.4) or (6.7) the conditions are for $a \geq 0$

$$Q_{j-\frac{1}{2}} - \frac{Q_{j+\frac{1}{2}}}{r} \leq 2\,, \tag{8.12a}$$

and for $a < 0$

$$Q_{j+\frac{1}{2}} - \frac{Q_{j-\frac{1}{2}}}{r^+} \leq 2 \tag{8.12b}$$

When the fully discrete case is considered (using backward Euler time differencing), the cell-edge based limiters conform to the same restrictions as the cell-centered types, but the semi-discrete form given above does have implications for some limiters discussed later in the chapter. Later some conditions are given with regard to certain fully discrete cases.

Remark 25 *Davis [189] discusses less restrictive limiters based on Lax-Wendroff type time centering. These limits are stated for* $a > 0$

$$Q_j < \frac{2}{\sigma a} . \tag{8.13a}$$

and

$$\frac{Q_j}{r} < \frac{2}{1 - \sigma a} . \tag{8.13b}$$

with analogous limits for $a < 0$.

One caveat applies to the strict use of conditions such as (8.11a)-(8.12b): the TVD conditions should be derived for each scheme from those stated in Theorem 6. An example of this principle at work is the derivation of appropriate limiters defined in Chapter 6 for parabolic PCT schemes. The resulting conditions for the limiters are identical to those above, but the right-hand sides of the inequalities are multiplied by 4/3. A simple example of this is the minbar limiter, (8.17), which produces a TVD scheme, but the proof of this requires a slight modification of the usual proofs (i.e., dropping the assumption that the Q functions are positive or equal to zero for all r).

Two Argument Limiters

Roe [176] and Sweby [132] defined their schemes (and limiters) to be upwind biased in nature. The stencil for the limiters was centered about a cell-edge and the cell-edge upwind from that. The typical assumptions regarding the positivity of the Q functions leads to the TVD region defined by Sweby. The boundary of this region is given by

$$Q_{TVD} = \mathrm{m}(2, 2r) . \tag{8.14}$$

It is bounded below by the x-axis. The TVD region using this assumption is shown in Fig. 8.2a. If the assumption regarding positivity is dropped then the region is bounded by

$$Q^1_{TVD} = \mathrm{m}(1, r) , \text{ and } Q^2_{TVD} = \mathrm{m}(-1, -r) . \tag{8.15}$$

This region is shown in Fig. 8.2b. Figure 8.2b differs from previous presentations in its recognition of limiters that can differ in sign (an example of which is the minbar

limiter).

As discussed in detail below, the limiters provide a second order [illegible] they make the scheme a convex [illegible] range of two (or more) [illegible] second order methods. For a upwind-biased cell-edge limiter, this means the limited [illegible] range of the Beam-Warming (Q_{BW}) and Lax-Wendroff (Q_{LW}) schemes. The second-order region of the plane for positive definite limiters [illegible] the region bounded by the minmod and superbee limiters (see below [illegible] more general cell-edge limiters given in Fig. 8.2, the second-order region is the upper bound [illegible] shaded region for $r > 0$ and the entire shaded region for r [illegible] 0.

For cell-edge limiters applied to second-order [illegible] and centered time discretizations (forward Euler or forward Euler with [illegible] Lax-Wendroff type correction), the regions given in Fig. 8.2 also apply. For fully [illegible] schemes based around the same methodology, the TVD region meets the [illegible] but the second order region of the plane remains the same, thus for practical purposes yielding the same sort of limiters. This point becomes significant when considering the ULTIMATE limiter in Section 8.3.5.

Several of the more common limiters are the basic "minmod" limiter [130]

$$Q_{\text{mm}}(1, r) = \text{mm}(1, r) , \tag{8.16a}$$

van Leer's limiter [119]

$$Q_{\text{vl}}(1, r) = \frac{r + |r|}{1 + |r|} . \tag{8.16b}$$

the centered limiter [120]

$$Q_c(1, r) = \text{mm}\left[2, 2r, \frac{1}{2}(1 + r)\right] , \tag{8.16c}$$

and Roe's superbee limiter

$$Q_{SB}(1, r) = \max[0, \min(2r, 1), \min(1, 2r)] \tag{8.16d}$$

Another form of limiter is used with the TVD type schemes. This limiter is called the "minbar" limiter and it returns the argument with the smallest absolute value. It can be written symbolically

$$\text{mb}(a, b) = \begin{cases} a & \text{if } |a| = \min(|a|, |b|) \\ b & \text{otherwise} \end{cases} \tag{8.17}$$

and in the ENO schemes the difference stencil grows in the direction of the smaller argument. Figure 8.3a shows the behavior of this limiter in the ratio form in the context of a second-order TVD scheme. In Fig. [illegible] second order TVD region

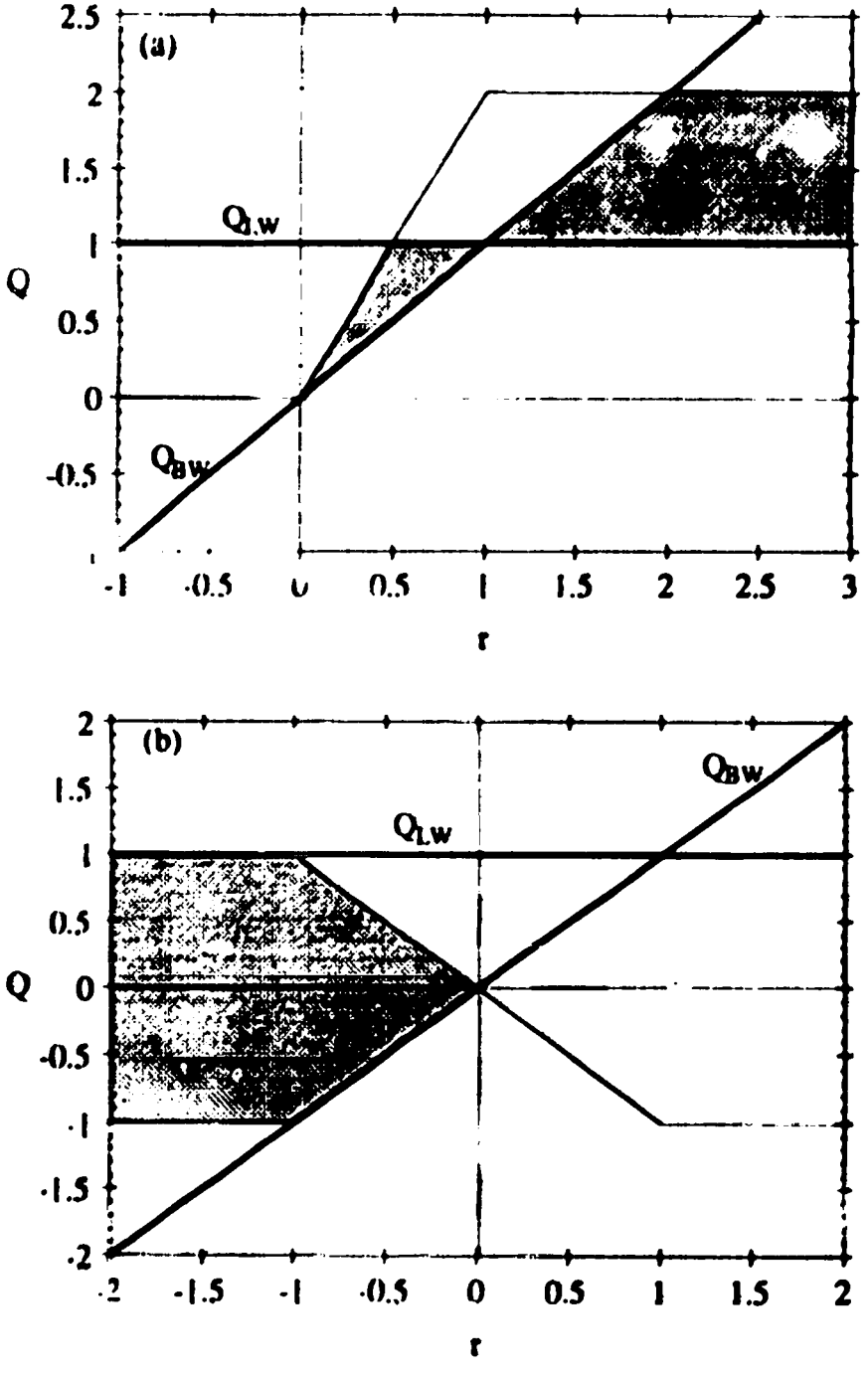


Figure 8.2: The second order TVD regions are shown in the shaded regions of these figures. The other lines show the limits of the TVD region for an explicit time differencing. Figure 8.2b gives the TVD regions assuming Q is positive definite. This agrees with the presentation given by Sweby. Figure 8.2a shows the TVD region assuming Q is not positive definite. The second-order TVD region includes the lines $Q = r$ for $0 \le r < 1$ and $Q = 1$ for $r \ge 1$. The lines denoted by Q_{LW} and Q_{BW} correspond to the Lax-Wendroff and Beam-Warming methods. The regions lying between these curves are second-order accurate. The other "thin" lines outline the TVD regions. In Fig. 8.2a this is the r-axis for $r > 0$. For Fig. 8.2b this is the line $Q = r$ for $0 < r < 1$ and $Q = 1$ for $r > 1$.

is shown as outlined by Q_l and Q_{SB}. Figure 8.4b shows the behavior of Q_c and Q_{ul} with respect to r.

For the initial presentation of this analysis, for example, the determination of the limiter at cell-edge $j + \frac{1}{2}$ if the signal velocity $a > 0$ then the gradient at $j - \frac{1}{2}$ is compared with the gradient at $j + \frac{1}{2}$, otherwise the gradient at $j + \frac{3}{2}$ is used for comparison. This scheme is (6.7) with

$$\tilde{s}_{j+\frac{1}{2}} = Q_{j+\frac{1}{2}} \frac{\Delta_{j+\frac{1}{2}} u}{\Delta_{j+\frac{1}{2}} x} . \tag{8.18}$$

The question of accuracy of limited schemes of this nature was addressed by Sweby in [132]. The schemes of this nature could be viewed as convex averages of the Lax-Wendroff and Beam-Warming schemes. These schemes are defined by the use of certain gradient ratios defined in a linear manner. The second-order TVD region is a set of the regions bounded by these two schemes and the conditions defining TVD schemes. A secondary effect of this is that the limiters thus become convex but nonlinear averages of the sample gradients. Two third-order upwind methods can also be incorporated into this framework. One is based on cell averages and the other is point value based [190] (see Chapter 9). These schemes are defined for the upwind-biased TVD schemes with gradients written in ratio form as

$$\frac{s}{s_{j+\frac{1}{2}}} = \frac{3}{4} + \frac{r}{4} , \tag{8.19a}$$

for the point-value form and

$$\frac{s}{s_{j+\frac{1}{2}}} = \frac{2}{3} + \frac{r}{3} , \tag{8.19b}$$

for the cell-average form. Figure 8.3b shows the region defined by these limiters in the second-order TVD region.

The use of these identical limiters has not been limited to schemes of this type. The HOG scheme described by Colella in [123] and Osher in [179] and the modified flux TVD scheme of Harten [130, 61] successfully use these same limiters. The polynomial interpolation for this scheme is given by (8.7). The limiters are not biased with the direction of the flow, and the limiters stencil is invariant. These schemes determine a value for the gradient which is cell-centered and is based on sample gradients taken at the cell edges. Analysis of conditions resulting in TVD limiters yields identical results as the upwind-biased limiter applied to a TVD Lax-Wendroff scheme as discussed later. In fact, for a scalar wave equation these two schemes give identical results with identical limiters. This does not generalize to nonlinear equations.

The accuracy of these schemes is second order in the L_1 norm, but the limiters make the resulting scheme a convex average of a second-order upwind scheme and the corresponding anti-upwind interpolated scheme. The first scheme produces results of

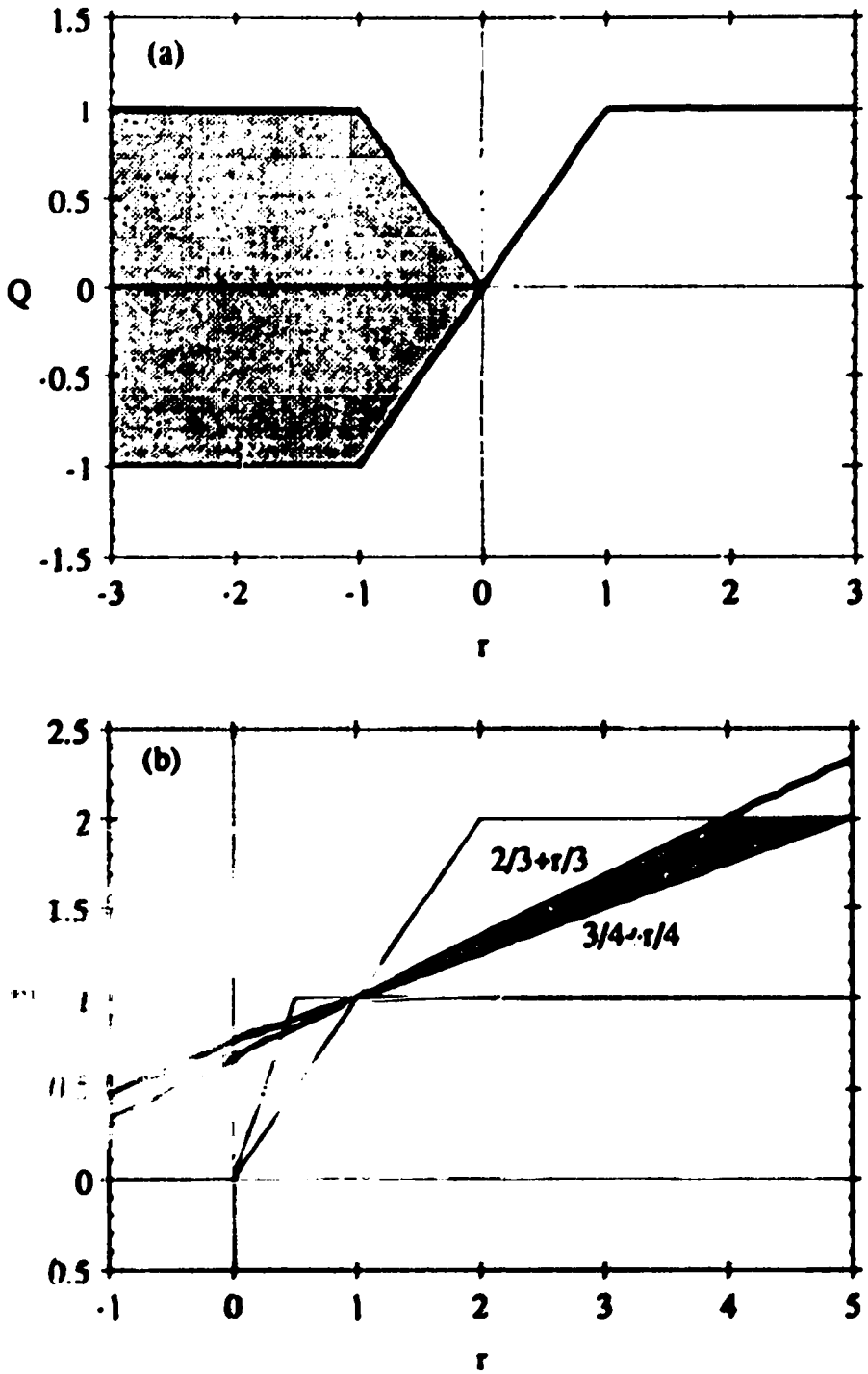


Figure 8.3: This shows the minbar limiter. It is interesting to note that for an upwind-biased cell-edge scheme this limiter gives a Beam-Warming scheme for $|r| \leq 1$ and a Lax-Wendroff method for $|r| \geq 1$. Figure 8.3b shows the third-order region of the plane.

relatively good quality, while the second scheme produces poor results (saved by the Riemann solver), but the limiter provides exceptional results improved in all respects. The relation of the linear difference schemes to the high order method is akin to the relation of Sweby and Roe's scheme and Lax-Wendroff or Beam-Warming schemes.

Before going further, several other two argument limiters should be introduced. The form used to define the minmod and superbee limiters are specific cases of a family of schemes defined by

$$Q_n = \max[0, \min(n, r), \min(1, n\,r)] \ , 1 \le n \le 2 \ . \tag{8.20}$$

For $n = 1$ this reduces to the minmod limiter and for $n = 2$ it is the superbee limiter. The above caveat also applies to this limiter because for some possible schemes the above definition can be extended. Figure 8.5a shows the behavior of Q_n for $n = 1.5$.

Osher and Chakravarthy [180] introduced a limiter

$$Q_{OC} = m(1, n\,r) \ \text{ or } \ m(n, r) \ 1 \le n \le 2 \ . \tag{8.21}$$

which does not share the symmetry condition with the other limiters (unless $n = 1$) and thus must be used with caution. This can be seen in Fig. 8.5b for $n = 2$ and each of the two forms given above. The first of these two choices makes sense from the standpoint that in a upwind-biased cell-edge limiter it would choose the centrally differenced gradient. The results presented in [181, 132] show the effects of this lack of symmetry. This limiter may still be used if applied carefully in algorithm construction. Nevertheless, these limiters find widespread use in a number of schemes and produce quality results in spite of their less desirable qualities.

Uniformly nonoscillatory schemes [64] use a limited second derivative to correct the first derivative estimate to give uniform second-order accuracy in all error norms. The price paid is the loss of the TVD property; however, these schemes are designed not to create any new extrema not in the initial data (for linear problems). For the polynomial form (8.7), the sample gradients used are cell-edge centered. The UNO scheme makes an estimate of the second derivative at the cell-edges and corrects the value of the cell-edge first derivative to the cell center. I define

$$d_j = \frac{s_{j+\frac{1}{2}} - s_{j-\frac{1}{2}}}{\Delta x} \ , \tag{8.22}$$

as the second derivative computed from the first derivatives $s_{j\pm\frac{1}{2}}$, and compute an estimate for $d_{j+\frac{1}{2}}$ with

$$d_{j+\frac{1}{2}} = m(d_j, d_{j+1}) \ \text{ or } \ m(d_j, d_{j+1}) \ . \tag{8.23}$$

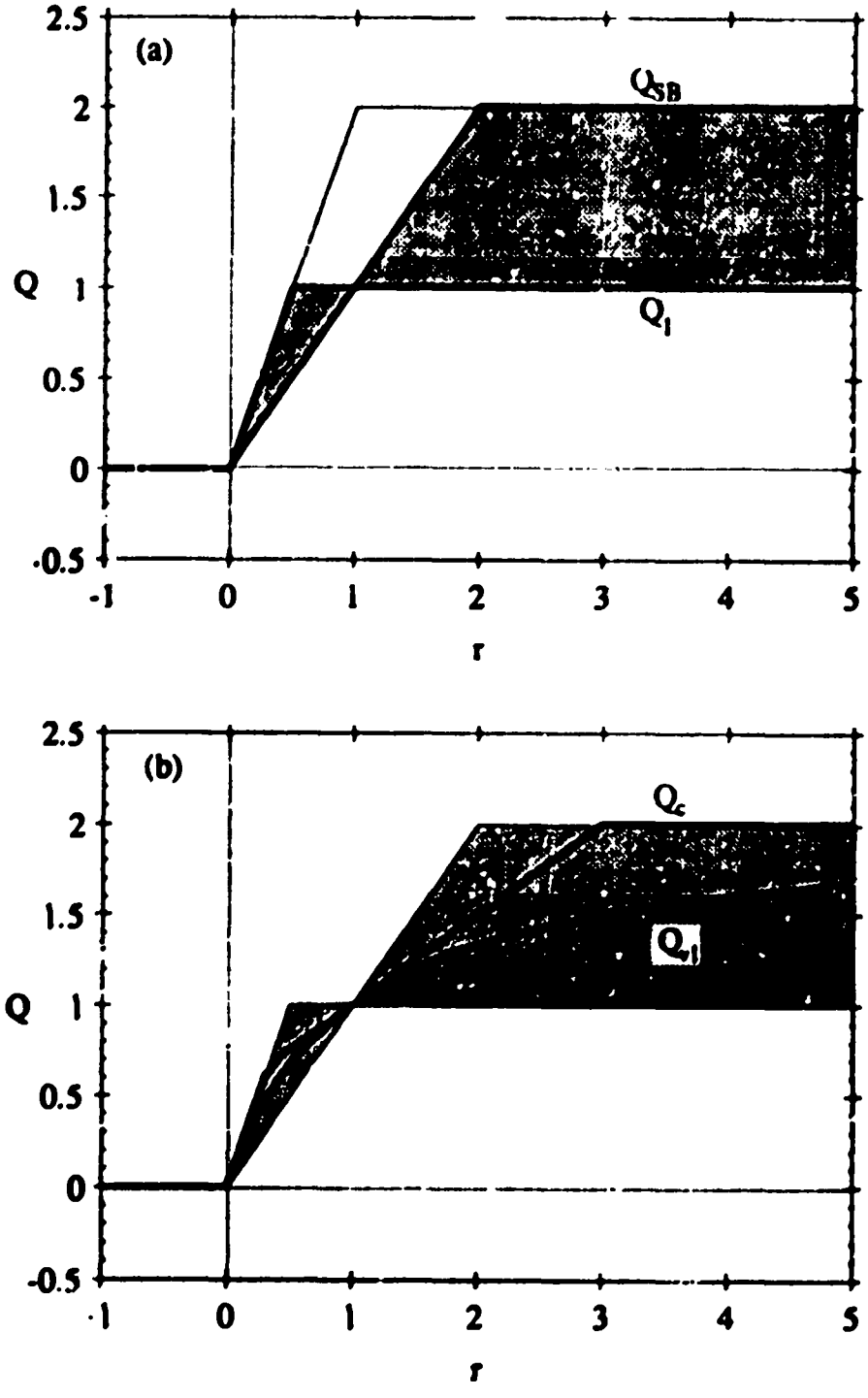


Figure 8.4: Figure 8.4a shows the minmod and superbee limiters. The minmod limiter gives the lower boundary and the superbee limiter gives the upper boundary of the second-order TVD region. In Fig. 8.4b, van Leer's and the centered limiter are given.

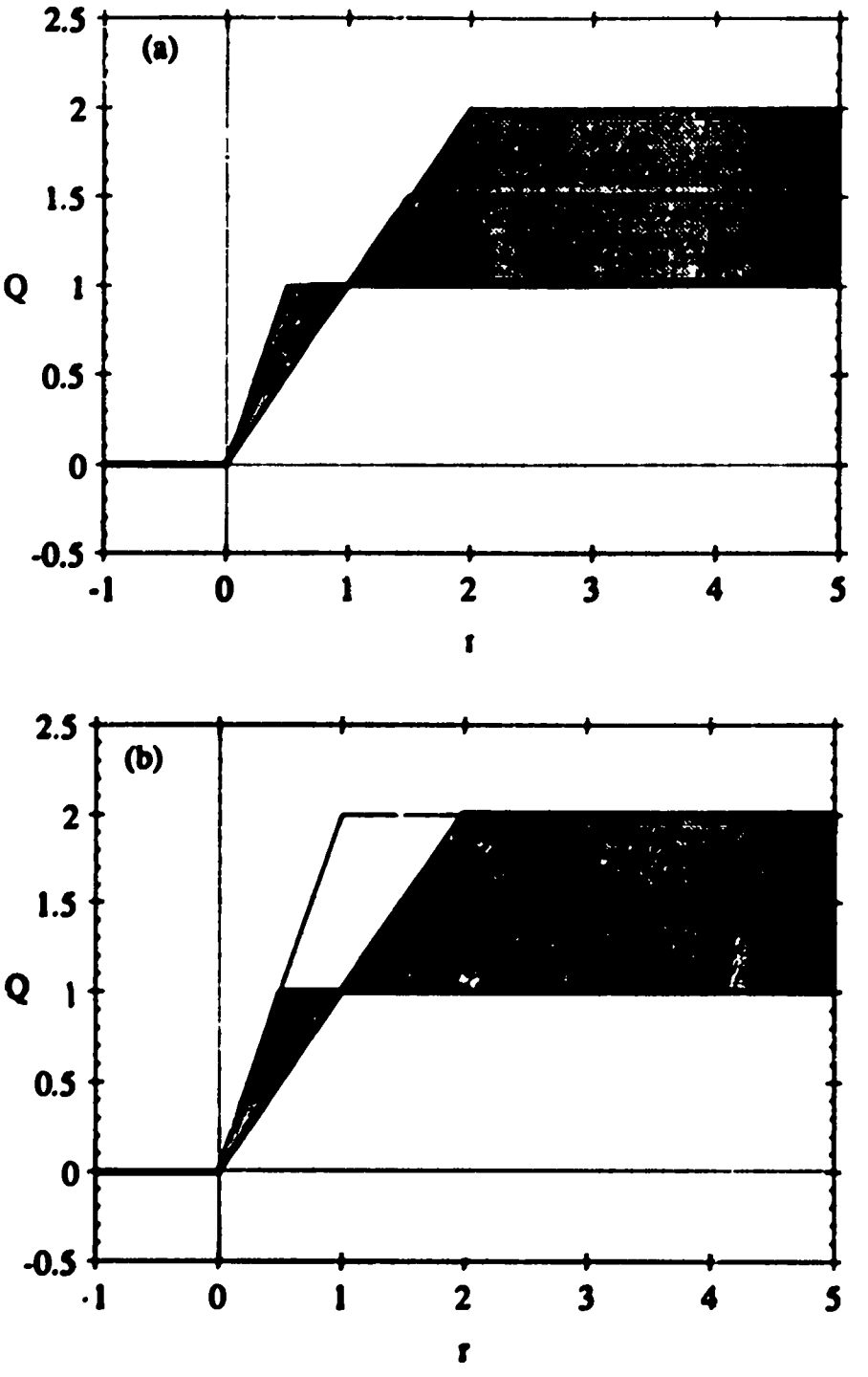


Figure 8.5: Figure 8.5a shows the limiter, Q_n, for $n = 1.5$. The plot shown by Fig. 8.5b looks similar to Fig. 8.3a, the difference is that the upper boundary of the second-order TVD region is given by one of the two limiters ($Q_{OC} = \mathrm{m}(1, 2r)$)for $r < 1$ and by the other ($Q_{OC} = \mathrm{m}(2, r)$) for $r > 1$.

I correct the first derivative estimates

$$s_{j-\frac{1}{2}} = s_{j-\frac{1}{2}} + \frac{\Delta x}{2} d_{j-\frac{1}{2}} \tag{8.24a}$$

and

$$s_{j+\frac{1}{2}} = s_{j+\frac{1}{2}} - \frac{\Delta x}{2} d_{j+\frac{1}{2}} \,. \tag{8.24b}$$

and limit these modified gradients in a normal fashion. The performance of this scheme on test problems is generally exceptional. This approach works for the modified flux TVD method and its related HOG counterpart. Suresh and Huynh have studied some interesting variant of the above UNO-type schemes [191].

The upwind-biased cell-edge limiter uses two argument limiters as well, but the proper definition of UNO requires some modification. Discussion of this is deferred to the next section.

The compressive limiters are necessary for computing contact discontinuities because of their tendency to diffuse. Less compressive limiters are recommended for shocks because of a shock's self-sharpening nature.

Q_{vl} does not have the usual form, but checking its functionality shows what its effect is. This can also be viewed as a modified harmonic mean. This connection is explored at length in Section 8.3.4.

Three Argument Limiters

As the discussion in the previous section would indicate, the two argument TVD limiters are relatively simple to analyze and take a number of forms. The three argument limiters are more difficult to analyze, but I follow the same general methodology.

Several limiters of this class have already been given in Chapter 7. To present these limiters in as compact a form as possible, the nomenclature used in Section 7.2 is used. Thus the following variables are defined:

$$r^- = \frac{\Delta_{j-\frac{1}{2}} u}{\Delta_{j+\frac{1}{2}} u} \,, \quad r^+ = \frac{\Delta_{j+\frac{3}{2}} u}{\Delta_{j+\frac{1}{2}} u} \,, \tag{8.25}$$

and the function $Q_{j+\frac{1}{2}}\left(s_{j-\frac{1}{2}}, s_{j+\frac{1}{2}}, s_{j+\frac{3}{2}}\right)$ can be rewritten as $Q_{j+\frac{1}{2}}\left(r^-, 1, r^+\right) s_{j+\frac{1}{2}}$. The term $s_{j+\frac{1}{2}}$ has the same definition as before. Some of the limiters of this class have been reported by Roe [131] and Yee [134]. Some example of these limiters are

$$Q_1\left(r^-, 1, r^+\right) = \mathrm{m}\left(r^-, 1, r^+\right) \,, \tag{8.26a}$$

$$Q_2\left(r^-, 1, r^+\right) = \mathrm{m}\left[2r^-, 2, 2r^+, \frac{1}{2}\left(r^- + r^+\right)\right] \,, \tag{8.26b}$$

and

$$Q_1'\left(r^-,1,r^+\right) = m\left(r^-,1\right) + m\left(1,r^+\right) - 1\,. \tag{8.26c}$$

Figure 8.6 shows these limiters. Limiters of the form of Q_1' are not recommended because of their behavior near discontinuities and extrema. Roe [131] noted this behavior by defining this type of function as a "separable Q function." This represents a simple manner of extending two argument limiters to the three argument case. Examples of this philosophy are extensions of the superbee and van Leer's limiter

$$Q_{vl}\left(r^-,1,r^+\right) = \frac{|r^-| + r^-}{1 + r^-} + \frac{|r^+| + r^+}{1 + r^+} - 1\,, \tag{8.26d}$$

and

$$\begin{aligned} Q_{sb}\left(r^-,1,r^+\right) = {} & \max\left[0,\min\left(1,2r^-\right),\min\left(2,r^-\right)\right] \\ & + \max\left[0,\min\left(1,2r^+\right),\min\left(2,r^+\right)\right] - 1\,. \end{aligned} \tag{8.26e}$$

If a function being limited is smooth and monotone over range of the three arguments being limited, no problem occurs because a monotone variation is assumed here. Problems occur when the data shows more structure. This is evident through Fig. 8.7, which shows that both of the above limiters are not TVD although their behavior in practice may be acceptable on most initial data.

At this point, several topics are in need of discussion. As before with the two argument limiters, accuracy of the approximation is important, and as before some criteria such as symmetry needs to be met. These allow us to create new limiters with desirable qualities.

The topic of accuracy can be addressed quite simply, as part of the answer comes from the previous analysis of upwind-biased limiters for the TVD Lax-Wendroff schemes. The three argument limiters (I am considering those centered about a cell edge) are a convex average of the Lax-Wendroff and Beam-Warming methods, but also include an anti-Beam-Warming-type scheme where the stencil is taken to be opposite of upwind. Although the result of the limiter is a convex average of these schemes, it is second-order accurate. The stability of schemes such as FCT or symmetric TVD show the power of limiters to offset the effects of using anti-upwind data. This statement is somewhat misleading as anti-upwind data is dangerous at extrema and discontinuities and the limiters discussed here would choose data from elsewhere in the stencil at these points.

As noted with Fig. 8.2, the TVD regions for the three argument limiters can be visualized by projecting the regions shown in the plot in an additional coordinate direction.

The concept of symmetry in these limiters needs to be different than with the two argument case. Common sense dictates that the limiter should be symmetric about

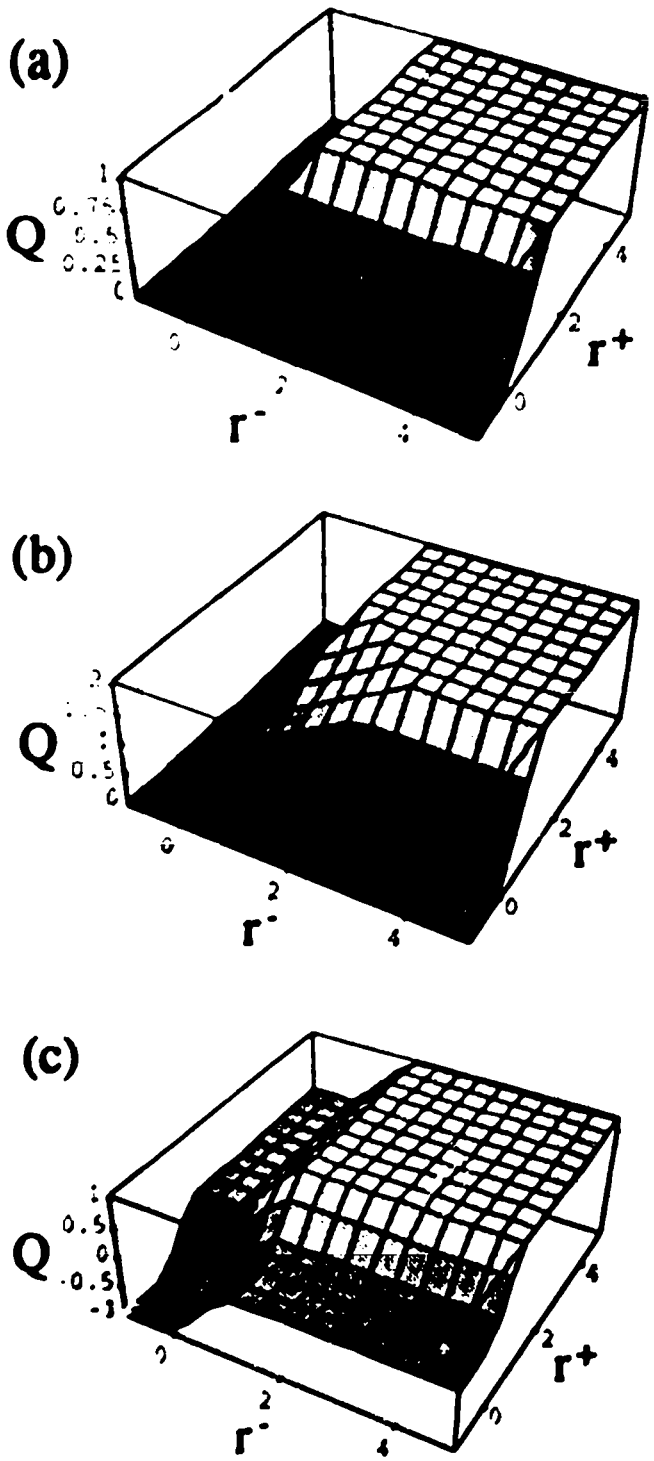


Figure 8.6: Three of the three argument limiters are shown here. These are the minmod limiter (Q_1), the centered limiter (Q_c), and a modified minmod limiter ((Q_1')). The modified minmod limiter does not give TVD results because of its form and subsequent behavior when $r^{\pm} < 0$. The other two limiter are TVD for second-order symmetric type schemes.

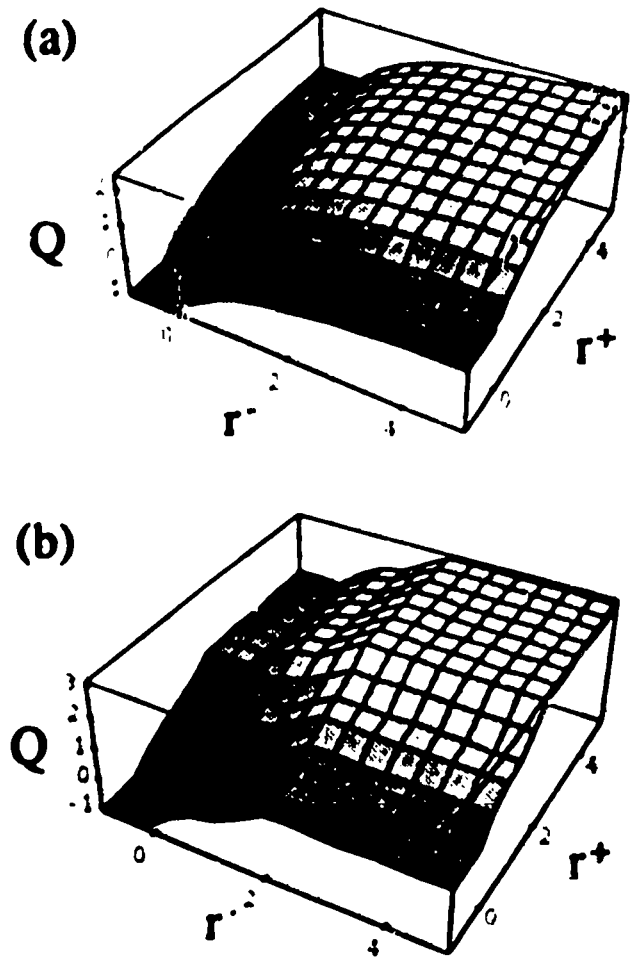


Figure 8.7: Both of these limiters use the design philosophy of the modified minmod scheme. Figure 8.7a uses van Leer's limiter and Fig. 8.7b uses the superbee limiter. Both are not TVD for $r^{\pm} < 0$, but also are not TVD should $r^{\pm}$ grow sufficiently large with both being greater than 1.

the central value in the stencil, i.e.,

$$Q\left(r^-, 1, r^+\right) = Q\left(r^+, 1, r^-\right) . \tag{8.27}$$

Inspection reveals that this is indeed the case for the limiters given above. The property of homogeneity is also important and is kept by the above limiters. The same caveat concerning limiters and specific difference schemes made in the previous section applies to the three argument limiters.

Before moving on, several limiters can be introduced that meet the above stated criteria. One limiter that quickly comes to mind is an extension of the minbar limiter, (8.17),

$$mt(a,b,c) = \begin{cases} a & \text{if } |a| = \inf\left(|a|, |b|, |c|\right) \\ b & \text{if } |b| = \inf\left(|a|, |b|, |c|\right) \\ c & \text{otherwise} \end{cases} . \tag{8.28}$$

Figure 8.8 shows this limiter behavior for different values of r^- and r^+. A general class of limiters extending two argument limiters to three arguments can be written

$$Q^3 = \min\left[Q^2\left(1, r^-\right), Q^2\left(1, r^+\right)\right] , \tag{8.29}$$

where Q^2 could be any two argument limiters like those discussed in the previous section. Two examples of this design principle are given in Fig. 8.9 (using van Leer's and the centered two argument limiters). This limiter does not share some of the poor characteristics of the separable limiters shown above. In several cases, the results from this limiter reduce to other limiters discussed above. For instance, the basic three argument minmod limiter can be found from the above combination of two argument minmod limiters.

A second group of limiters, which have their basis on the above-stated symmetry property, are natural outgrowths of several of the two argument TVD limiters. Examples of this design are

$$Q_c = \max\left[0, \min\left(2, 2r^-, 2r^+, \frac{1}{2}\left(1 + r^-\right), \frac{1}{2}\left(1 + r^+\right)\right)\right] , \tag{8.30a}$$

$$Q_{sb} = \max\left[0, \min\left(2, r^-, r^+\right), \min\left(1, 2r^-, 2r^+\right)\right] , \tag{8.30b}$$

and

$$Q_{vl} = \frac{|r^-| + |r^+| + r^- + r^+}{2 + |r^-| + |r^+|} . \tag{8.30c}$$

The limiters satisfy the TVD requirements for the symmetric TVD scheme and perform quite well in practice. These are shown in Fig. 8.10, which demonstrates their ability to produce symmetric TVD limiters.

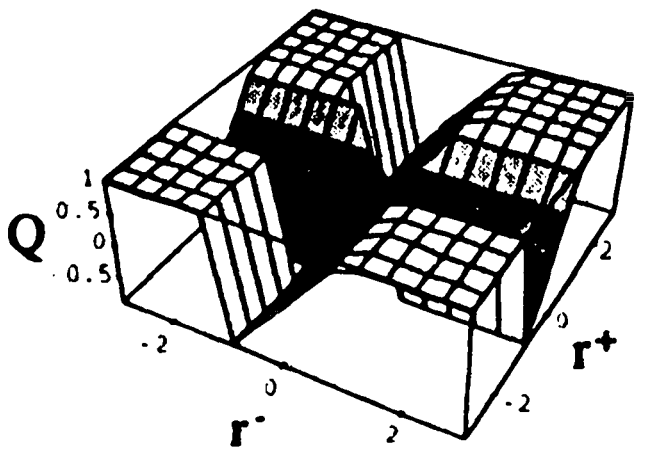


Figure 8.8: The three argument analog to the minbar limiter is shown here.

As discussed in Chapter 6, the concept of UNO schemes can be generalized to the three argument limiters. This is done in the following manner. The cell-edged stencil for the limiters requires that gradients one full cell distant from cell edge be used in the limiting process. As with the two argument implementation of an UNO scheme, these gradients are corrected. To do this, cell-edged estimates for the second derivatives are needed, as defined by (8.23). The gradients used in the limiter are then corrected with a first-order correction based on these second derivatives. The cell-edge gradient on the cell edge where the limiter is defined is already second-order and needs no correction. These corrections are

$$\hat{s}_{j-\frac{1}{2}} = s_{j-\frac{1}{2}} + \Delta x d_{j-\frac{1}{2}} \tag{8.31a}$$

and

$$\hat{s}_{j+\frac{1}{2}} = s_{j+\frac{1}{2}} - \Delta x d_{j+\frac{1}{2}} \,. \tag{8.31b}$$

As noted in the previous section, the upwind-biased limiters cannot use the UNO description given in the previous section. The cell-edge-based definition given in the previous paragraph is the proper basis to begin from and the generalization to the upwind-biased limiters is natural.

The methods introduced as being symmetric TVD schemes are differentiated by their flux limiters which are centered in support about the cell edges. The other methods like those introduced by Sweby and Roe are upwind biased in the support for their limiters. Both methods however are closely related to the Lax Wendroff method. The symmetric schemes have been favorably viewed because of their lower operation count and an increased convergence rate [166].

In considering the performance of these schemes, six test problems are completed: two for the scalar wave equation, one for Burgers' equation, and three for the Euler

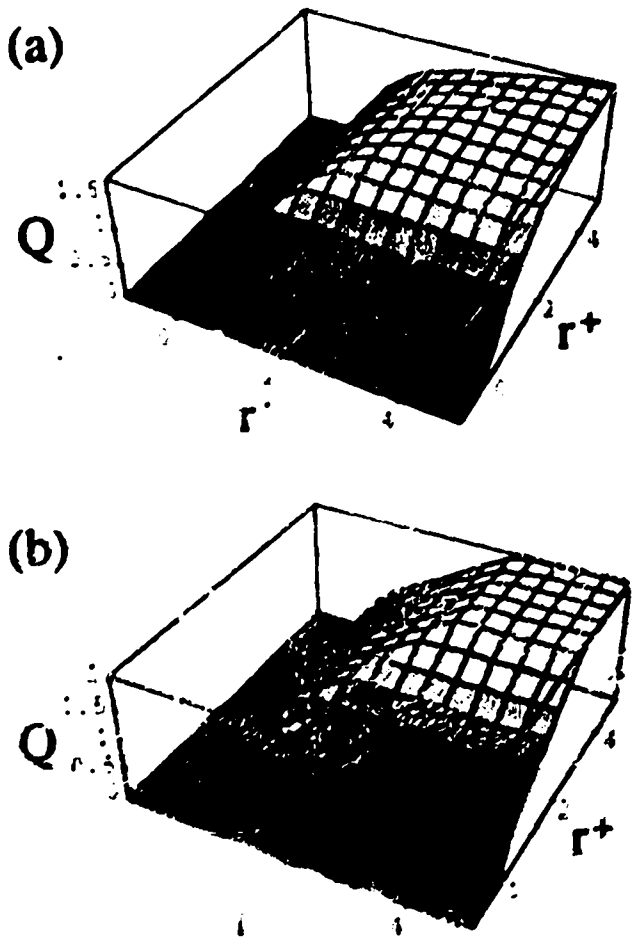


Figure 5.9. Here a different methodology is used to create three argument limiters. The resulting limiters are TVD and do not suffer from the same difficulties as the modified minbar type of limiter. The two base limiters used here are van Leer's and the centered limiters. In practice any TVD two argument limiter can be used in this context.

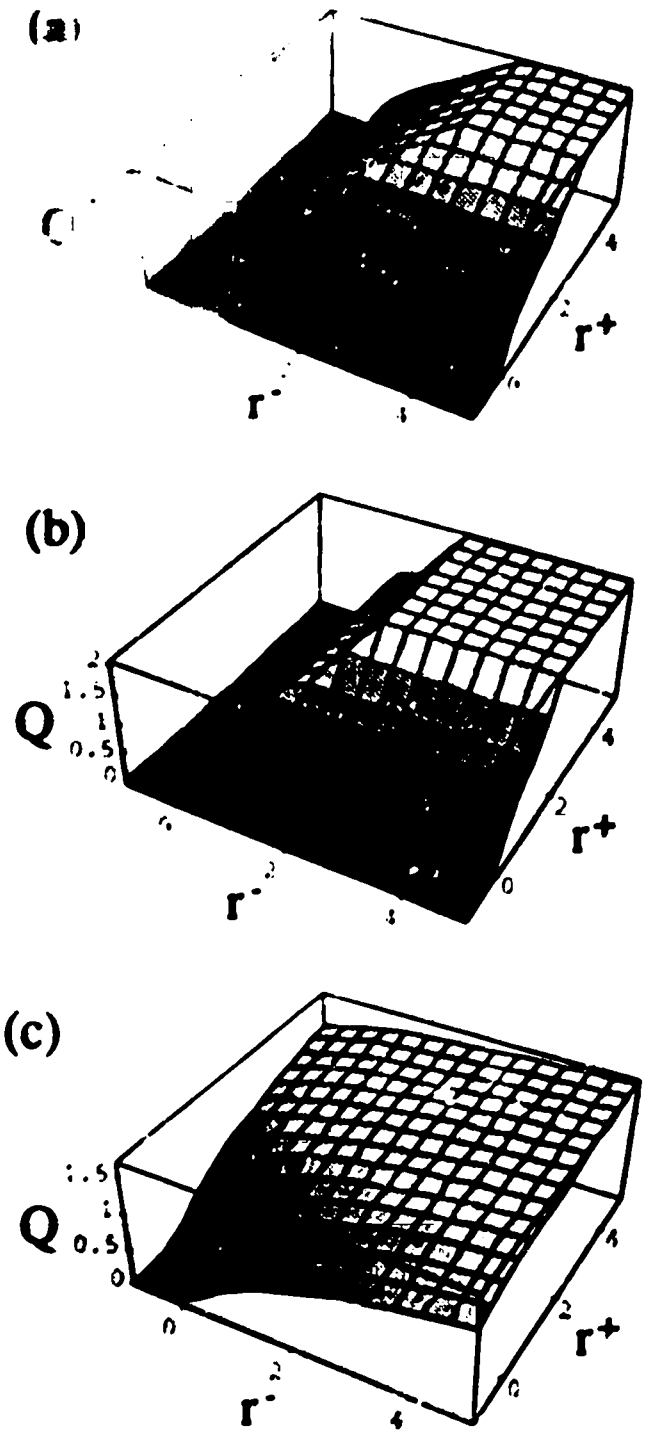


Figure 8.10: The limiters shown here use the symmetry property discussed in the text. The limiter shown in Fig. 8.10a is analogous to the centered limiter while Fig. 8.10b is analogous to the superbee limiter. Both are second order and TVD. Figure 8.10c gives a van Leer type limiter, which is not TVD but works quite well in practice.

Table 8.1: Order of accuracy in several norms for the schemes solving Burgers' equation.

Scheme	L_1	L_2	L_∞
Symmetric ($t = 0.2$)	1.83	1.58	1.19
Upwind ($t = 0.2$)	1.90	1.65	1.28
Symmetric ($t = 1.0$)	1.48	1.19	0.78
Upwind ($t = 1.0$)	1.41	1.14	0.74

equations. The two problems for the scalar wave equation are the advection of a square wave and of a "teepee" function across a periodic domain. Each test runs for 300 time steps with a Courant-Friedrichs-Lewy (CFL) number of $\frac{1}{2}$. The Burgers' equation problem is simply a $\sin(x)$ initial condition on a periodic domain with length of 2π. The three Euler equation problems are Sod's problem [41], Lax's problem [55], and a blast wave problem [44]. The combination of these problems highlights the strengths and weaknesses of these algorithms. Both algorithms always use the limiter denoted by Q_2 in the previous section for all problems except the Burgers' equation problem where Q_1 is used.

Figure 8.11 shows the solutions to the scalar wave equation. The symmetric scheme obviously provides lower resolution in both cases. The difference is also fairly great in terms of both peak preservation as well as signal width. The symmetric scheme also has problems with signal shape as it is somewhat distorted. A notable feature of the upwind-biased scheme is that for the scalar wave equation the solution is identical to that obtained by the modified flux TVD scheme if the same limiters are used. This can be explained by the support of the limiter used and the resulting interpolation on the upwind side of each cell interface. For nonlinear problems this does not hold.

In Table 8.1, the rates of convergence are given for Burgers' equation. When the solution is smooth, the upwind method is evidently superior in every error norm. After a shock forms, the symmetric scheme is slightly more convergent; however, for all test cases (up to 1000 grid cells) the actual error is lower for the upwind scheme. In addition, as time progresses after $t = 1.0$, the upwind scheme recovers its initially higher rate of convergence.

The solutions for the Euler equations echo the results with the previous three problems. Across the board, the resolution afforded by the upwind scheme is superior. The major flow structures: shocks, rarefactions, and contact discontinuities are all noticeably better resolved with the upwind method. The results from Sod's problem demonstrate this to some degree. In Fig. 8.13, each of the features are sharper with

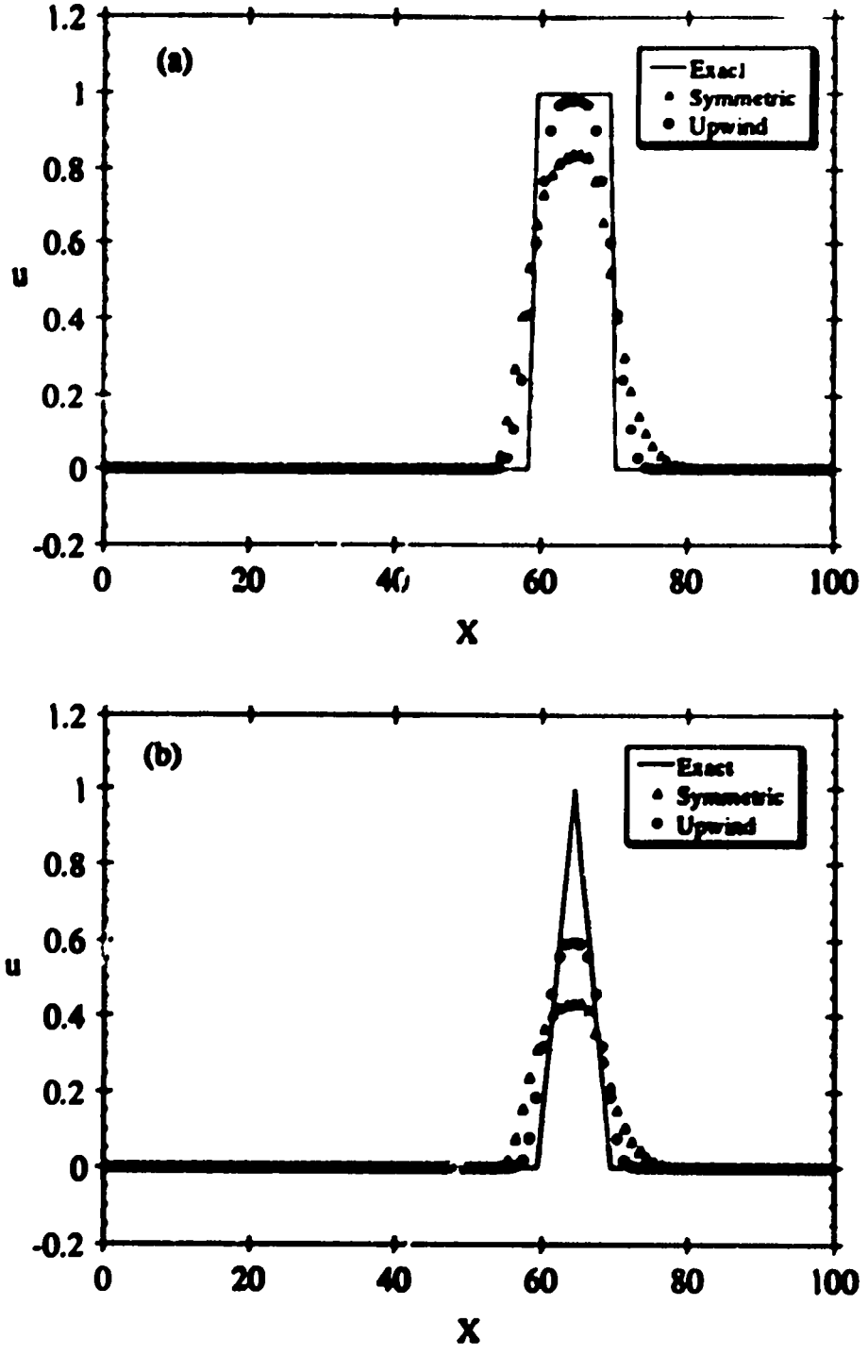


Figure 8.11: The solution of the scalar wave equation by both these methods is shown for two test problems. In both cases, the upwind method provides superior performance.

the upwind method. This is probably most noticeable at the contact discontinuity. In Fig. 8.12, the noted behaviour for the contact discontinuity and shock are clearly shown. Also evident from this figure is the symmetry problems exhibited by the symmetric scheme. The shape of the density peak is more consistent with the exact solution with the upwind-biased method.

The blast wave problem (see Fig. 8.14) accentuates each of these issues. This is particularly true with respect to the right density peak which is significantly closer to the converged solution with the upwind method. Two other key features of the solution are the degree of "fill-in" between the peaks and the contact discontinuity to the left of the left density peak. The fill-in regions are both smeared nearly equally, but the shape of the upwind computed solution is better. The left-most contact discontinuity is much more smeared by the symmetric scheme.

The results of the previous paragraphs show conclusively that the upwind scheme produces results of higher resolution when compared with the symmetric scheme. This raises the issue of cause. These schemes are second-order accurate when the solution is smooth. The limiters are based on minimum principles, and increasing their support lowers the value returned by the function. The subsequent "flattening" of the slope is akin to increasing the numerical viscosity of the scheme thus lowering the accuracy.

Interpreted on a more physical basis, the upwind scheme takes data from a more physically meaningful location on the grid. The support for the limiter can be perceived to affect the solution at that point, whereas the symmetric limiters are centered by taking both upwind and antiupwind data. Both arguments lead to a conclusion that if resolution is of primary concern, the limiter should have as small a support as possible in order to limit its induced viscosity. This of course should be within the limitations of providing physically meaningful oscillation-free (or nearly so) results.

Appendix E provides the results of using both two and three argument limiters without limiting for each term.

Artificial Compression

Often, it is important to choose the limiter used by the nature of the problem. For fields that are linearly degenerate, the problem of numerical diffusion is severe. In the solution of systems of equations this manifests itself as severe smearing of contact discontinuities. A number of schemes have been developed to deal with this problem [183, 122, 110, 137, 192, 193]. One such scheme is artificial compression, which can be applied to TVD limiters. The form is

$$Q_j = (1 + \omega_j \theta_j) Q_j \,. \tag{8.32a}$$

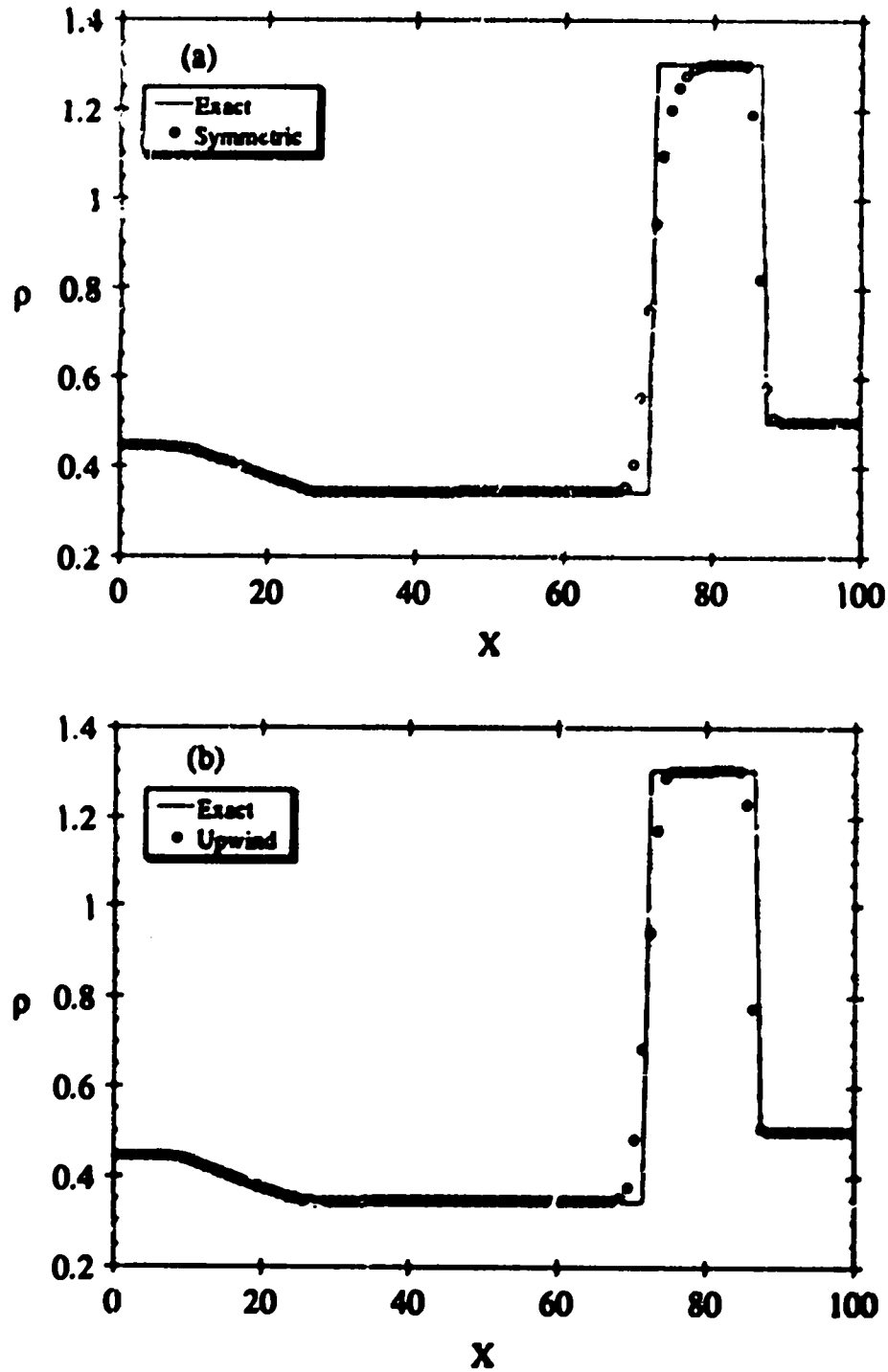


Figure 8.12: The solution to Lax's problem highlights the resolution of both shocks and contact discontinuities as well as the symmetry properties of the solution: methods.

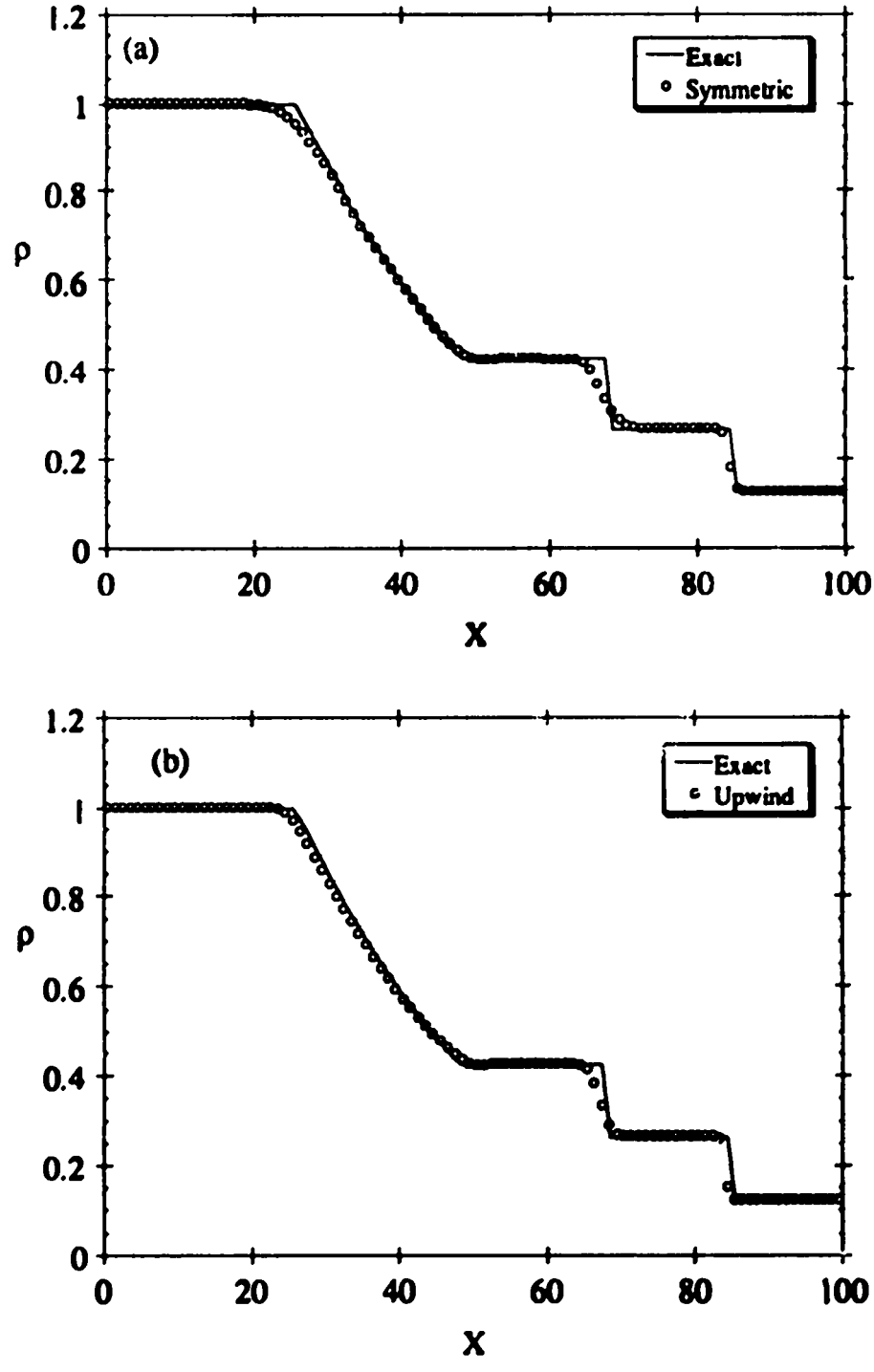


Figure 8.13: The solution to Sod's problem by both methods shows the improved resolution given by the upwind-biased scheme.

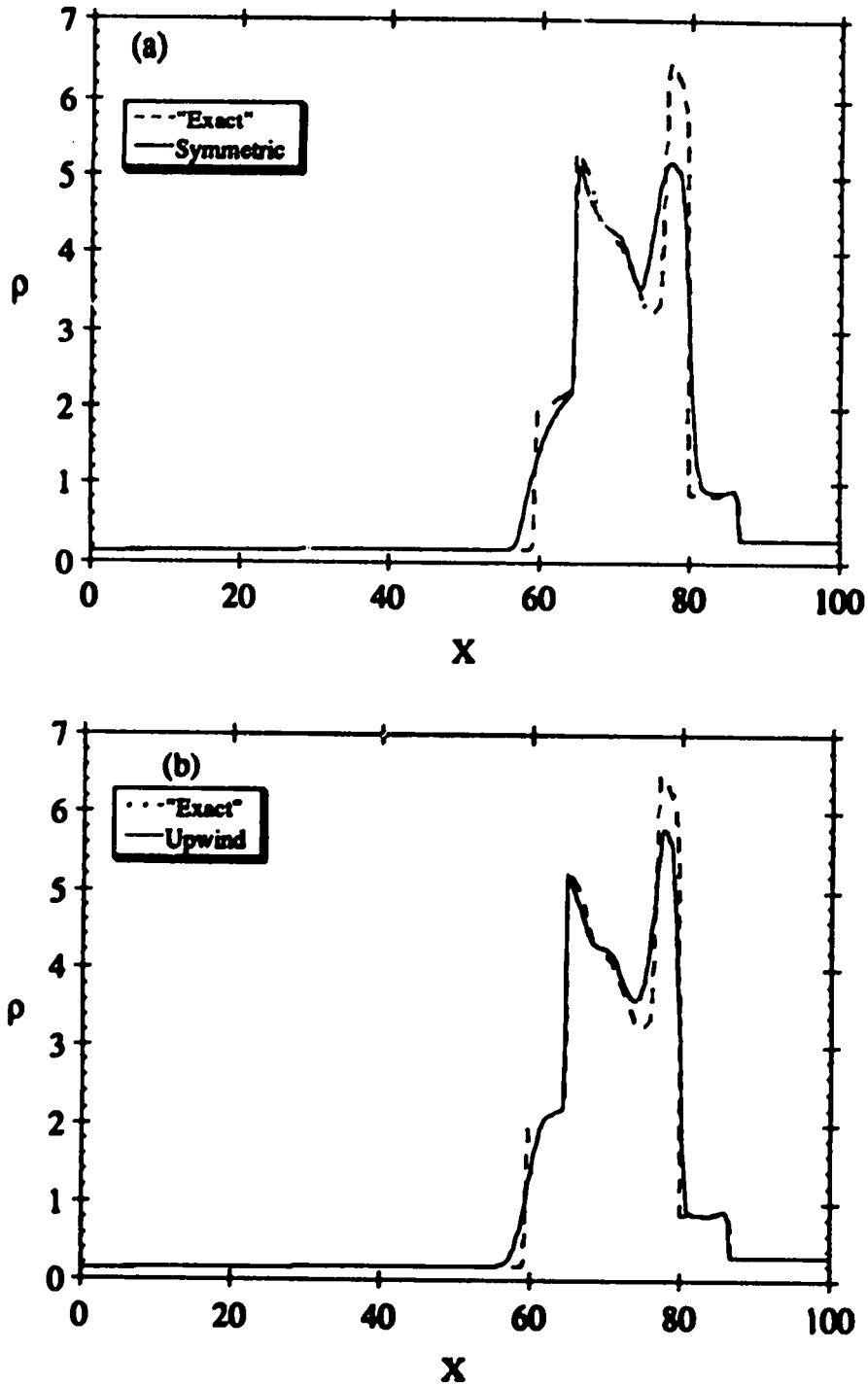


Figure 8.14: In the blast wave problem, the deficiencies of both methods are most clearly shown. The difficulty of the problem is due to the large amount of structure confined to a small physical space.

where the discontinuity detector, θ_j is defined as

$$\theta_j \equiv \frac{\left|\Delta_{j+\frac{1}{2}} u - \Delta_{j-\frac{1}{2}} u\right|}{\left|\Delta_{j+\frac{1}{2}} u\right| + \left|\Delta_{j-\frac{1}{2}} u\right|} = \frac{|1 - r|}{1 + |r|} , \tag{8.32b}$$

and the argument, ω_j, is chosen to give the best results. Figure 8.3.3a shows how θ varies with r. This applies compression to the method (makes the local slope steeper). If the field is genuinely nonlinear, then the limiter should not be so compressive in nature.

An effective form for ω_j in transient problems was introduced in [101]. This was used with the superbee limiter under the stipulation that the resulting scheme remained TVD after the application of artificial compression. This application was not second order in the sense of the definition given in the previous sections. With the superbee limiter the form is

$$\omega_j = \min\left(|\nu_j|, 1 - |\nu_j|\right) , \tag{8.33}$$

where ν_j is the local CFL number. A more general form can be found that produces TVD results (for common TVD schemes like those presented in Section 8.3.3). This form is

$$\omega_j = 2 - \xi + \min\left(|\nu_j|, 1 - |\nu_j|\right) , \tag{8.34}$$

where $\xi = \max\left[Q(r)\right]$ $r \in \Re$.

For the case of three argument limiters, artificial compression is generally not applied. The same general form used above can be used after several modifications. The discontinuity detector is applied to two sets of gradients when choosing the maximum value is

$$\theta_{j+\frac{1}{2}} = \max\left(\theta_j, \theta_{j+1}\right) , \tag{8.35}$$

and ω is computed at the cell-edges. The behavior of θ for the cell-edged three argument case is shown in Fig. 8.3.3b. The effectiveness of this approach is discussed in Section 8.4.

A large degree of caution should be exercised when using artificial compression or similar schemes. The type of limiter used and the compression involved appears to affect solutions solved for long time periods on periodic domains [159]. The more compressive algorithms can give completely erroneous results while less compressive ones converge to the correct solution. In steady-state solutions the less compressive limiters normally give more convergent solutions. This is the likely outcome of increased dissipation present in the algorithms. In this example, the FCT method of Boris and Book produced exceedingly poor results that can probably be attributed to the amount of compression in the algorithm.

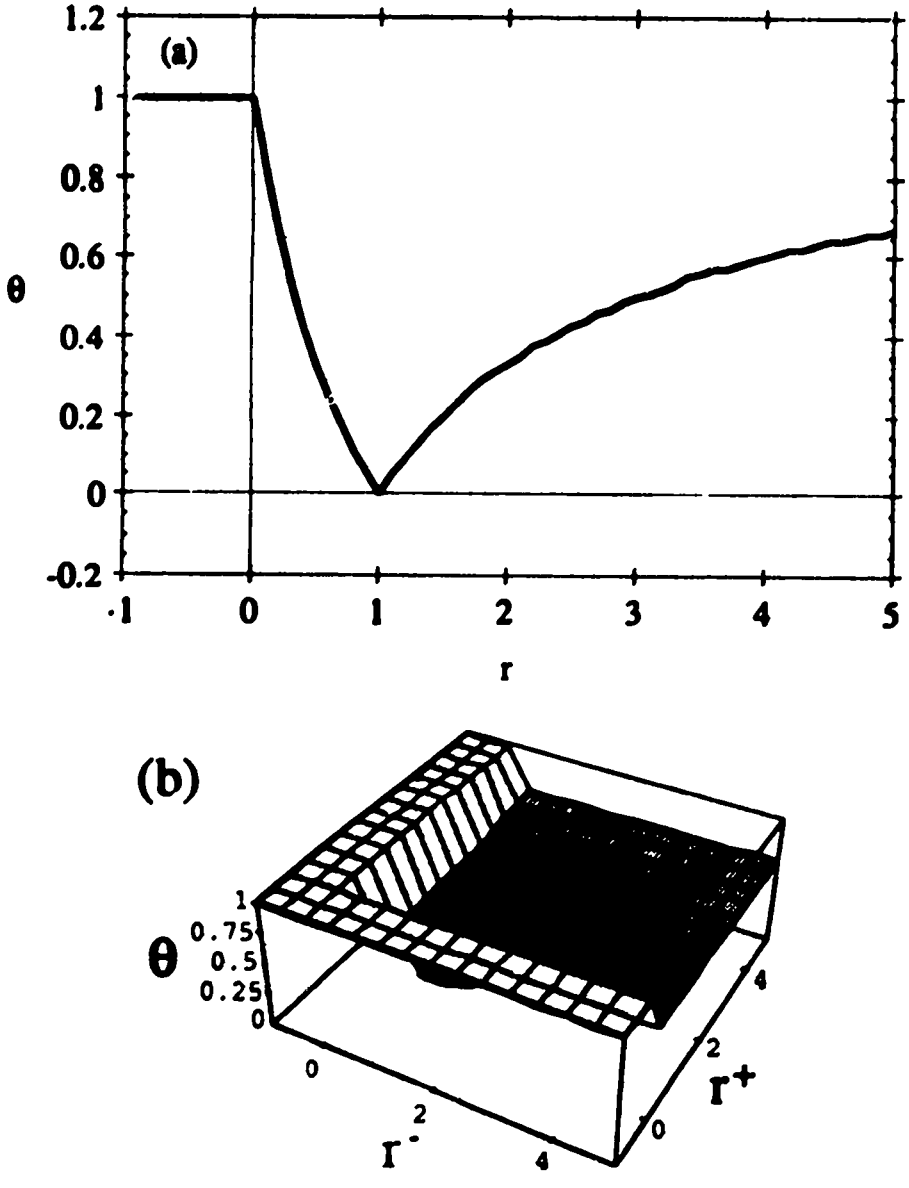


Figure 8.15: Here the behavior of the discontinuity detector in the artificial compression algorithm is shown for use with both two and three argument limiters.

8.3.4 Nearly TVD Limiters

The previous sections concentrated on limiters that meet TVD criteria for the commonly used TVD schemes. By its nature, maintaining a TVD solution requires that the solution reduce to first-order accuracy at extrema. For long transients or those involving a great number of time steps, the impact of this is profound. In virtually every commonly reported solution, peaks are clipped and the solution is diffused. It is not reasonable to expect this to change as these are intrinsic to numerical approximation, but the degree to which these errors occur should be improved. Where the solution is not diffused and the front remain sharp, often smooth transitions are unphysically sharpened by the action of the limiter. Thus the currently used limiters are not always equal to the task.

To attempt improvement on some of the above-mentioned problems it may be useful to relax the requirement that a scheme produce a TVD solution. One way of doing this is to use a different definition for variation control of the scheme. This approach has been taken by Shu [169] in the total variation bounded (TVB) schemes. I have also looked into a more general view of limiters as a nonlinear average of the sample gradients as a manner of approach to this problem. Other approaches employ ENO type discretizations and/or least squares methods [165].

TVB Limiters

Shu has developed TVB schemes as a uniformly high-order alternative to TVD schemes. The TVB property simply requires that

$$TV(u,t) \leq B \tag{8.36}$$

for some time $t > 0$. This requires that basic TVD limiter be modified to take advantage of this definition (TVD implies that a scheme is TVB). This modification requires that some estimate of the second derivative of the solution be made in an *a priori* manner. Higher order derivatives have to be estimated if higher than second-order schemes are needed. This quantity is defined by the symbol M. This estimate then modifies the gradients in the limiter that are not centered about the point being limited. The effect of this is to bias the limiter into choosing the higher-order centered gradient. This allows oscillations to form in the solution, but when they grow too large the nonlinear action of the limiter stops the growth. Although this has not been proved, it is believed that ENO schemes are TVB [65, 66].

The details of implementation can be divided into several distinct groups based on the type of limiter being used. For two argument limiters centered on the grid point the limiter must be divided into two pieces, each centered on the cell edge. Thus the limiter

$$\widetilde{\Delta_i u}^{TVD} = Q(1,r)\,\Delta_{i-\frac{1}{2}} u \tag{8.37a}$$

becomes

$$\widetilde{\Delta_j u}^{TVB} = \frac{1}{2}\left[Q(1, r+m)\,\Delta_{j-\frac{1}{2}} u + Q(1, r+m)\,\Delta_{j+\frac{1}{2}} u\right] , \tag{8.37b}$$

where $m = M\Delta x / s_{j-\frac{1}{2}}$ or $m = M\Delta x / s_{j+\frac{1}{2}}$ for the appropriate term in (8.37b). Examples of this limiter are shown in Fig. 8.16 for two values of $M\Delta x$. Here the definition of the limiter function Q has not changed from that given in Section 8.3.3, but its arguments have. The argument away from the cell edge where the limiter is centered has $M\Delta x$ added to it, thus the limiter is in most cases biased towards the selection of the argument it is centered on. A proof of the TVB nature of this limiter is given in [169].

Several approaches can be taken to implementing this methodology with cell-edged limiters. The method described above for cell-centered limiters can be used with slight modification. The upwind-biased cell-edge limiter is defined by

$$\tilde{s}^{TVB}_{j+\frac{1}{2}} = Q(1, r+m)\, s_{j+\frac{1}{2}} , \tag{8.38}$$

where m is defined as above and r is the ratio of the upwind gradient from cell-edge $j+\frac{1}{2}$ and $s_{j+\frac{1}{2}}$. For the centered cell-edge limiters, the approach follows the logical extension of the upwind-biased case. In this case a limiter is defined by

$$\tilde{s}^{TVB}_{j+\frac{1}{2}} = Q\left(r^- + m, 1, r^+ + m\right) s_{j+\frac{1}{2}} . \tag{8.39}$$

Figure 8.17 shows this limiter for two values of $M\Delta x$. On the plateau of the figures, the schemes are second-order accurate and, as shown, the sizes of the plateaus increase with $M\Delta x$.

Theorem 10 *The limiters given by (8.38) and (8.39) result in a TVB scheme if these limiters and the resulting numerical schemes are TVD with $m = 0$. The resulting schemes (those considered here) are uniformly second-order accurate.*

Proof. The proof is similar to the proof given in [169]. If the underlying numerical scheme is TVD, then the proof reduces to showing that the total variation is bounded by some constant at all time, $t > 0$. This is accomplished through the use of a modified flux

$$\tilde{f}^{TVB}_{j+\frac{1}{2}} = \tilde{f}^{TVD}_{j+\frac{1}{2}} + \dot{c}_{j+\frac{1}{2}} , \tag{8.40}$$

which is the sum of a TVD flux and a constant. If it can be shown this constant is bounded, then its sum is bounded, in turn leading to an upper bound on the total variation. The accuracy argument involves showing that the constant M in the limiter creates a bias that results in the selection of the high-order accuracy gradient centered at the limiters location. □

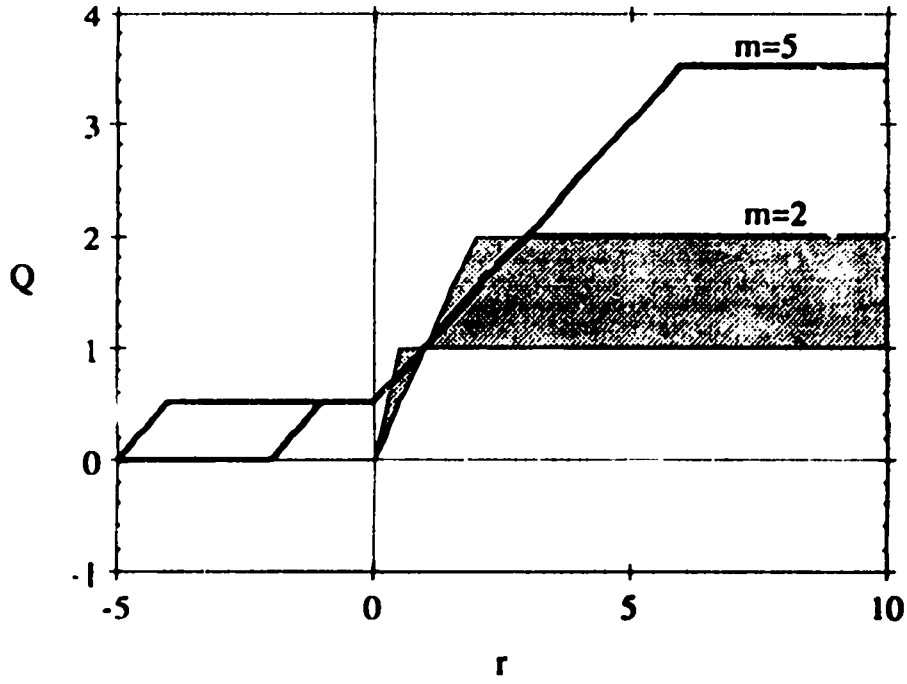


Figure 8.16: Two cases of the two argument TVB limiter are given here. The line that grows upward along the line $Q = \frac{1}{2}(1 + r)$ past $r = 3$ uses $M\Delta x = 5$ while the other line uses $m\Delta x = 2$. Both are always in the second-order region of the plane.

S–Limiters

One characteristic shared by the TVD limiters with the exception of the minbar limiters is setting the limited gradient to zero when the sign changes among the limiters arguments. The minbar limiter simply returns the argument that has the smaller absolute value, which may be opposite in sign to the function at that given point. This leads to a loss of accuracy at these points. As Tadmor [194] showed, the requirement for a scheme to be TVD (by Harten's definition) extrema must be clipped.

The limiters given in this section were designed to correct this problem. The essential feature of these schemes can be encapsulated in the following definition:

Definition 6 (S–limiters) *An S-limiter returns a value equal to some nonlinear average of its input arguments and has the same sign as the argument defined at the same location as the limiter.*

For example, in most cases this is some sort of gradient. The limited gradient has the same sign as the gradient at the location where the limiter is defined. For cell-edge-based algorithms, the changes in the reconstructive polynomial are minimal, but for cell-centered reconstructions some redefinition is required.

Starting from the scheme given by (8.7) and redefining it to meet the above-stated

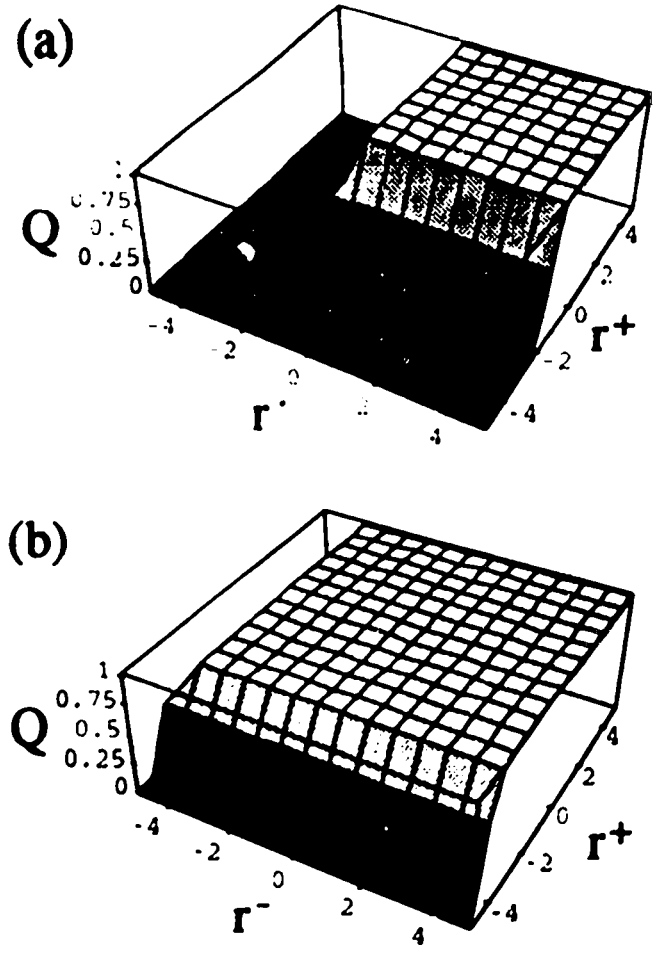


Figure 8.17: The three argument TVB limiter is shown here for $M\Delta x = 2$ and $M\Delta x = 5$. The larger value of $M\Delta x$ gives a larger "plateau" on the plot.

definition gives

$$P_j(x) = \bar{u}_j + \begin{cases} \text{sign}\left(\Delta_{j+\frac{1}{2}}u\right)\left|\widetilde{\Delta_j u}\right|\frac{(x-x_j)}{\Delta_j x} ; x \in \left[x_j, x_{j+\frac{1}{2}}\right] \\ \text{sign}\left(\Delta_{j-\frac{1}{2}}u\right)\left|\widetilde{\Delta_j u}\right|\frac{(x-x_j)}{\Delta_j x} ; x \in \left[x_{j-\frac{1}{2}}, x_j\right] \end{cases} . \tag{8.41a}$$

Here the gradient, $\widetilde{\Delta_j u}$, is redefined as

$$\widetilde{\Delta_j u} = S(1,r)\,\Delta_{j-\frac{1}{2}}u \text{ or } S(r,1)\,\Delta_{j+\frac{1}{2}}u\,. \tag{8.41b}$$

where the simplest example of the function S is

$$S_1(1,r) = \min(1,|r|)\ ; \tag{8.41c}$$

another example would be the centered limiter

$$S_c(1,r) = \min\left[2, 2|r|, \frac{1}{2}(1+|r|)\right]\ . \tag{8.41d}$$

These limiters are shown in Fig. 8.18. The term S_1 is a TVD limiter over its entire range, but S_c is not. The limiters can be logically extended to three arguments as before. One noteworthy point to raise with this reconstruction is that cell average of the reconstruction no longer equals the cell average $\bar{u}_j$ if sign $\left(\Delta_{j+\frac{1}{2}}u\right) \neq$ sign $\left(\Delta_{j-\frac{1}{2}}u\right)$. This subject is the topic of the next chapter.

In general, these limiters can be defined as above. They act as a multiplier on the cell-edge gradients modifying its magnitude but not its sign. This differs from the normal definition of limiters at points of extrema as noted above. The limiters are easily constructed from the definition of TVD limiters by removing the feature that sets the gradient to zero if the signs differ, and changing the reconstruction algorithm to one like the one shown above.

These limiters are not TVD unless the magnitude of $S(1,r) \leq 1$. Despite this, limiters of this nature perform well in practice (see Section 8.4) and have some advantages over the limiters constrained to be TVD. In test problems, the total variation was monitored and these limiters provide a TVD solution in practice. This may not hold true for all initial data.

Generalized Average Limiters

As noted in several sections above (8.3.3 and 8.3.3), limiters can be viewed as nonlinear averages of their arguments. In this section, this subject is explored further. As noted in Section 8.3.3, van Leer's limiter is a modified harmonic mean of its arguments. Another limiter was introduced in [159, 158], which has an interesting interpretation.

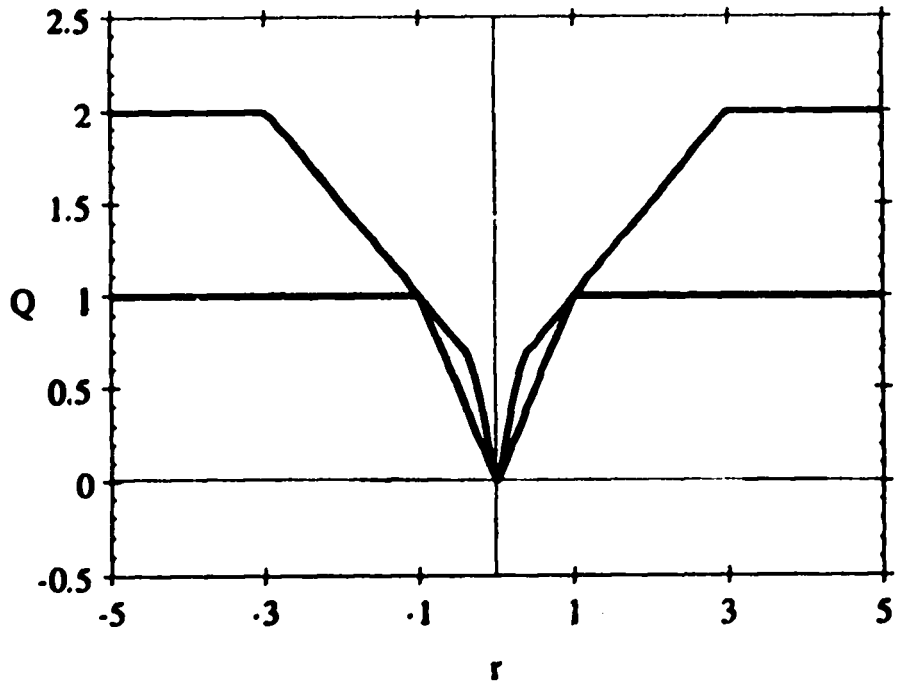


Figure 8.18: Two S-limiters are shown here. The upper of the two lines is for the centered limiter S_c while the lower is for S_1. S_1 is a TVD limiter.

This limiter is written

$$Q_{alb}(a,b) = \frac{(b^2+\delta^2)\,a + (a^2+\delta^2)\,b}{a^2+b^2+2\delta^2}\,, \tag{8.42}$$

where δ is a small positive bias. This bias is added to guard against clipping smooth extrema in the solution. Its role is similar to that of M in the TVB schemes. It should be chosen to be $|du/dx|$ [195] or $|du/dx|^{3/2}$ [159] from the smooth regions of the flow. Dropping δ and converting this to the normal form for analysis gives

$$Q_{alb}(1,r) = \frac{r^2+r}{1+r^2}\,. \tag{8.43}$$

This limiter can be written in an interesting form

$$Q_{alb}(a,b) = \frac{2ab}{a^2+b^2}\left[\frac{1}{2}(a+b)\right]\,. \tag{8.44}$$

In this form it has a nonlinear coefficient modifying the average of the input arguments. In [196], another form of this family of limiter was given (dropping the bias, δ) as

$$Q_{m-vl}(a,b) = \frac{2a^2b+2ab^2}{(|a|+|b|)^2} = \frac{4ab}{(|a|+|b|)^2}\left[\frac{1}{2}(a+b)\right]\,. \tag{8.45}$$

This limiter is more compressive than $Q_{a/b}$ and looks a great deal like the harmonic mean limiter. As $|a/b| \uparrow \infty$, $Q_{m-vl} \uparrow 2$. This limiter can also be written in ratio form as

$$Q_{m-vl}(1,r) = \frac{2r + 2r^2}{(1+|r|)^2} . \tag{8.46}$$

This limiter behaves exactly as Q_{vl} for $r \geq 0$, but for $r < 0$ it behaves differently (because it does not equal zero).

The noteworthy point is that both this limiter and van Leer's limiter can be written in a form that encompasses both of them as well as a much larger class of limiter. This form is

$$Q(a,b,n) = \frac{|a|^n b + |b|^n a}{|a|^n + |b|^n} , \tag{8.47}$$

or in a form suitable for analysis,

$$Q(1,r,n) = \frac{r + |r|^n}{1 + |r|^n} . \tag{8.48}$$

Limiters obtained for two values of n are given in Fig. 8.19a.

If one takes the limit as $n \uparrow \infty$, the minbar limiter is recovered, making it a limiting form of this family. For $n \neq 1$ or $n \neq \infty$ this limiter does not produce a TVD scheme in the numerical experiments, but the results are quite good. The comments contained in [159] are also of some importance when considering this limiter.

For more than two arguments, one can look to the suitable extensions of the definitions of harmonic mean and generalize to the power limiter above. For the three argument case this is

$$Q(a,b,c,n) = \frac{|ab|^n c + |ac|^n b + |bc|^n a}{|ab|^n + |ac|^n + |bc|^n} . \tag{8.49}$$

This limiter is shown for $n = 2$ (in ratio form) in Fig. 8.19b.

It is also interesting to investigate the results obtained with other nonlinear averages such as the geometric mean. The results obtained with this scheme are not TVD, but have some redeeming qualities.

8.3.5 The ULTIMATE Limiter

This limiter has received a great amount of attention in the literature recently. Leonard and coworkers [81, 82, 83] have presented this limiter in a series of papers. In another recent paper, this limiter was compared with other methods on shock tube problems [197]. The results showed that Leonard's limiter probably suffers from overcompression resulting in entropy violating solutions. In the following paragraphs, I discover where this characteristic arises in this method.

For this discussion, I do not use the system of nomenclature adopted by Leonard,

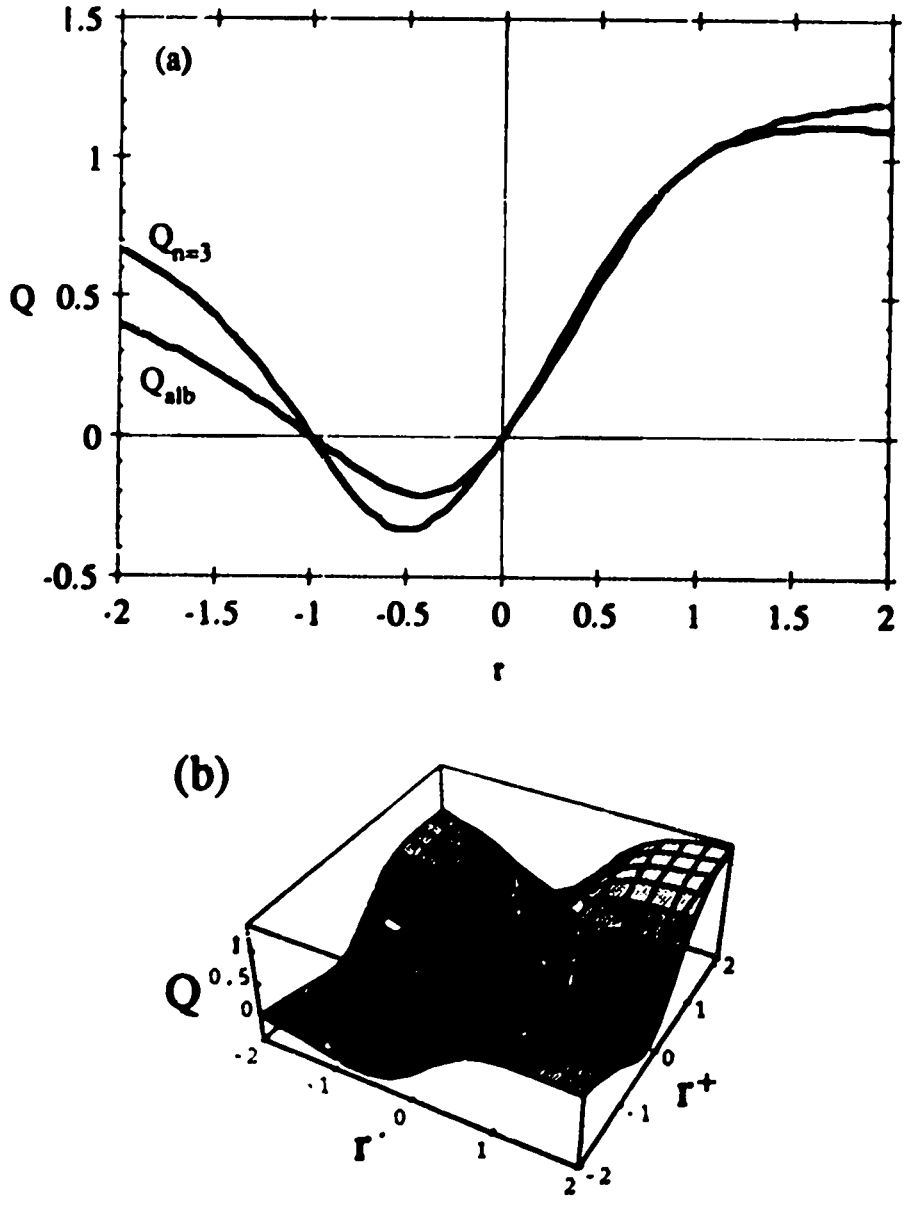


Figure 8.19: The generalized average limiter is shown in these figures. Figure 8.19a gives two examples of the two argument limiter for $n = 2$ and $n = 3$. Neither of these limiters is TVD. Figure 8.19b shows the $n = 2$ limiter for the three argument case.

but rather move his notation into the system adhered to earlier in this chapter. This should allow this limiter to be compared on a "level playing field." First, a short background is necessary. This method was developed in response to non-monotonic behavior of Leonard's QUICK[1] method in the presence of discontinuities. This method has been used extensively in engineering heat transfer type applications and represents the typical high-order scheme employed in those simulations. In this regard, Leonard's limiter is a great improvement, but its merits and shortcomings need more attention.

The normalized value diagram used by Leonard is not reviewed (one can refer to the above references), and simply move on to the presentation of the ULTIMATE limiter in my terms. Quite easily it can be shown that his limiter has the following form:

$$Q(r) = \min(\Delta u^a, C\,r, 2.) \ , \tag{8.50}$$

where $\Delta u^a = u^H - u^L$ is akin to the antidiffusive flux in the FCT method and C is some constant $\gg 1$. In his papers, Leonard uses $C = 200$. The value of u^H is determined by a linear high-order upwind method (like QUICK). This limiter is displayed in the usual fashion in Fig 8.20a. By including the QUICK differencing (the third-order point value scheme form Section 8.3.3) it can be seen that the region near the origin is not TVD for explicit time differencing.

Simple observation shows that the above limiter is not TVD for explicit temporal calculations unless $C = 2$ and u^a can be guaranteed to be within the bounds of a TVD limiter. When used with fully implicit time differencing or steady-state computations, the limiter is TVD. For $C > 2$, the limiter is no longer a convex average of second-order schemes and extremely compressive. This behavior is similar to that found with the FCT limiter. The saving grace is that the high-order upwind methods like QUICK are well-behaved approximations for hyperbolic conservation laws. It is highly likely that if other high-order centered approximations were used the limiters behavior would be far worse (much more compressive). In other words, the positive features of the underlying linear advection scheme mask some of the problems with the limiter.

A recent paper by Leonard [84] discusses the ULTIMATE limiter in transient problems. He suggests that $C = 2/\nu$. This yields a scheme which is nearly identical to the classic FCT without the diffusive first step. His results show that using a Lax-Wendroff or Beam-Warming type flux for the high-order flux with ULTIMATE yields poorer results than the better TVD limiters. Only when the third-order high-order flux is used are they better (not by much). Considering that the TVD schemes are essentially designed with Lax-Wendroff or Beam-Warming fluxes as the high-order fluxes those results are more applicable for limiter comparison.

[1] The QUICK method is a quadratic polynomial-based upwind method of third order accuracy.

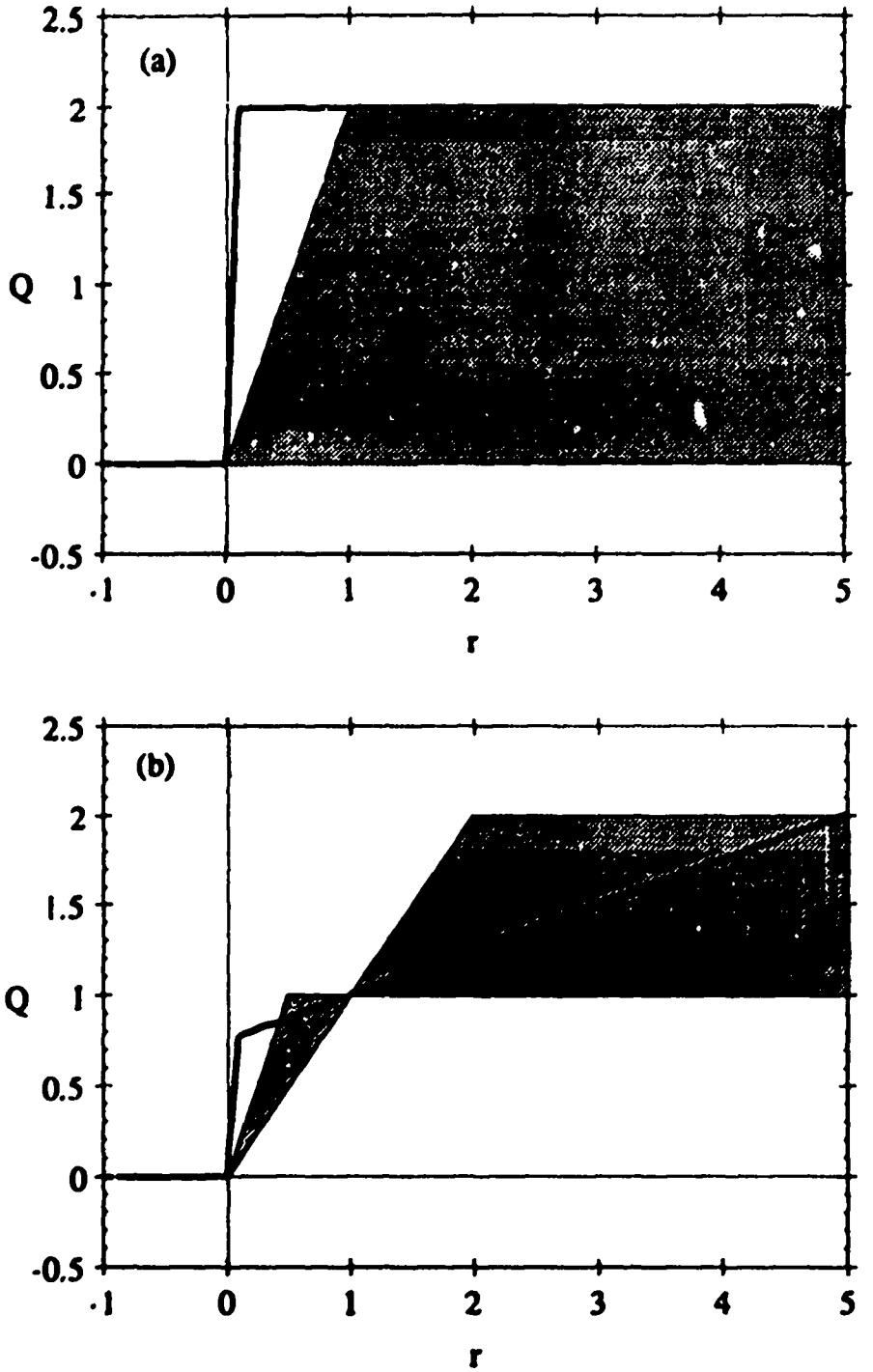


Figure 8.20: The ULTIMATE limiter is shown in this figure without the benefit of the high-order upwind flux. The basic limiter is not TVD for explicit time discretizations unless $C = 2$. The QUICK differencing is included in 8.20b. The region near the origin gives non-TVD results for explicit schemes.

8.4 Results

This section presents results for some of the limiters described in previous sections. The results are limited to the scalar wave equation and Burgers' equation. No attempt is made to present results for all the limiters given above, but the types of limiters introduced here are discussed with regard to their performance in relation to resolution and convergence. The solution of the Euler equations using these limiters could also yield useful information about the limiter. This is left for later investigations. With the exception of the FCT limiters, the basic numerical schemes used in the results is (8.7) for the two argument limiters and (6.7) for the three argument limiters. Table 8.2 shows a list of the limiters considered in the results and the abbreviations used in referring to them below.

The general characteristics of the test problems are given in Appendix A.

8.4.1 The Scalar Wave Equation

In this section using various limiters, the scalar wave equation is solved by the methods described in this chapter. Two initial conditions are used for the analysis: a square wave with a width of 10 cells and a $\sin^2 x$ wave (half of a period) of width of 25 cells. Both tests are conducted for 500 time steps with a CFL number of one-half. The advective velocity is taken to be unity.

The results for the TVD two and three argument limiters are given in Figs. 8.21–8.23. The results for most limiters are what can be expected. The three argument limiters make the resulting numerical scheme more diffusive, thus lowering the resolution of the solutions. One important point is the horrible performance of the SB3P limiter, which is not TVD. The SB2 limiter is also interesting because it seems to compress the $\sin^2 x$ wave into a square wave. This behavior is commonly seen with this limiter and warrants some warning. It is primarily caused by the limiter not being able to differentiate between a diffused square wave and the smooth $\sin^2 x$ wave. The limiter "recognizes" it as diffusion and compresses it. Various results regarding the resolution, accuracy, and numerical diffusion can be seen in Tables 8.3– 8.5. For the limiters of these categories, these tables show no surprises except in the case of the SB2 limiter. By the measure of numerical diffusion used here this limiter actually provides negative diffusion. This is not unstable because it is applied in a nonlinear fashion. Where positive diffusion is needed, the limiter supplies it. For the $\sin^2 x$ problem, the CENT2 and VL2 limiters are more accurate than SB2.

The results for artificial compression show that its effects are similar to that produced by the superbee limiters in both the two and three argument cases. Figure 8.24 shows that the artificial compression results in sharper profiles and increased resolution when compared with the normal minmod limiter. For the form of implementation used here, the resulting solution is not as compressed as with the superbee limiter.

The TVB solutions are shown in Figs. 8.25 and 8.27. The two argument TVB lim-

Table 8.2: Abbreviations for the methods used in this study.

Limiter	Equation	Abbreviation
Two Argument Minmod	(8.16a)	MM2
Two Argument van Leer	(8.16b)	VL2
Two Argument Centered	(8.16c)	CENT2
Two Argument Superbee	(8.16d)	SB2
Three Argument Minmod	(8.26a)	MM3
Three Argument Minmod Prime	(8.26c)	MM3P
Three Argument Superbee	(8.30b)	SB3
Three Argument Superbee Prime	(8.26c)	SB3P
Three Argument van Leer	(8.30c)	VL3
Three Argument Centered	(8.26b)	CENT3
Two Parameter Artificial Compression Minmod	(8.32b)	MM2A
Three Parameter Artificial Compression Minmod	(8.35)	MM3A
Two Argument Minmod TVB	(8.37b)	MM2TVB
Three Argument Minmod TVB	(8.51b)	MM3TVB
Signed Two Argument Minmod	(8.41c)	SMM2
Signed Two Argument Centered	(8.41d)	SCENT2
Signed Three Argument Minmod	(8.41c)	SMM3
Signed Three Argument Centered	(8.41d)	SCENT3
Two Argument van Albada	(8.43)	VA2
Three Argument van Albada	(8.49)	VA3
Two Argument van Albada with Bias	(8.43)	VA2B
Three Argument van Albada with Bias	(8.49)	VA3B

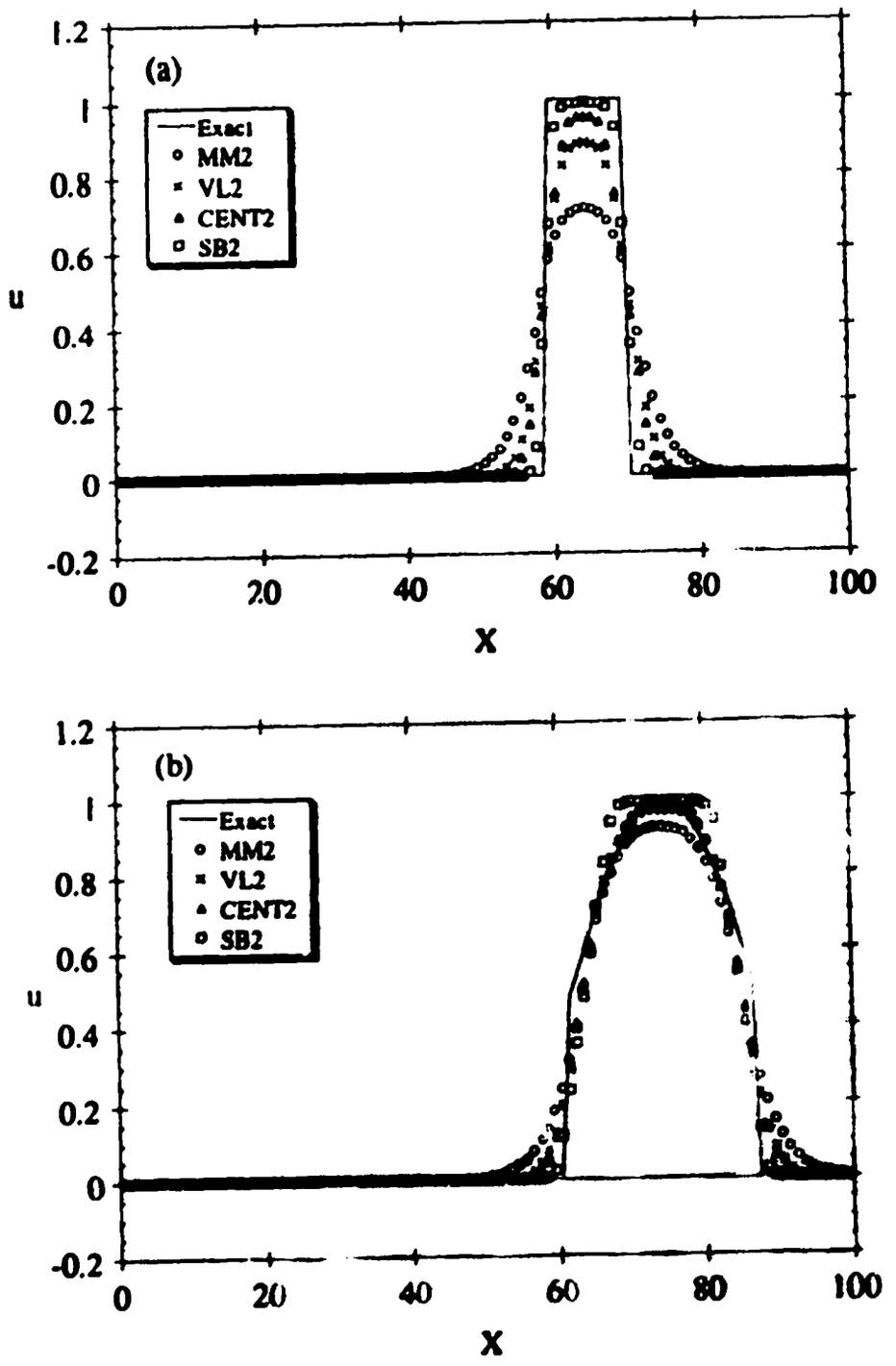


Figure 8.21: The scalar square and $\sin^2 x$ wave solutions using several two argument TVD limiters. Note that the SB2 limiter compresses the $\sin^2 x$ profile into a square wave.

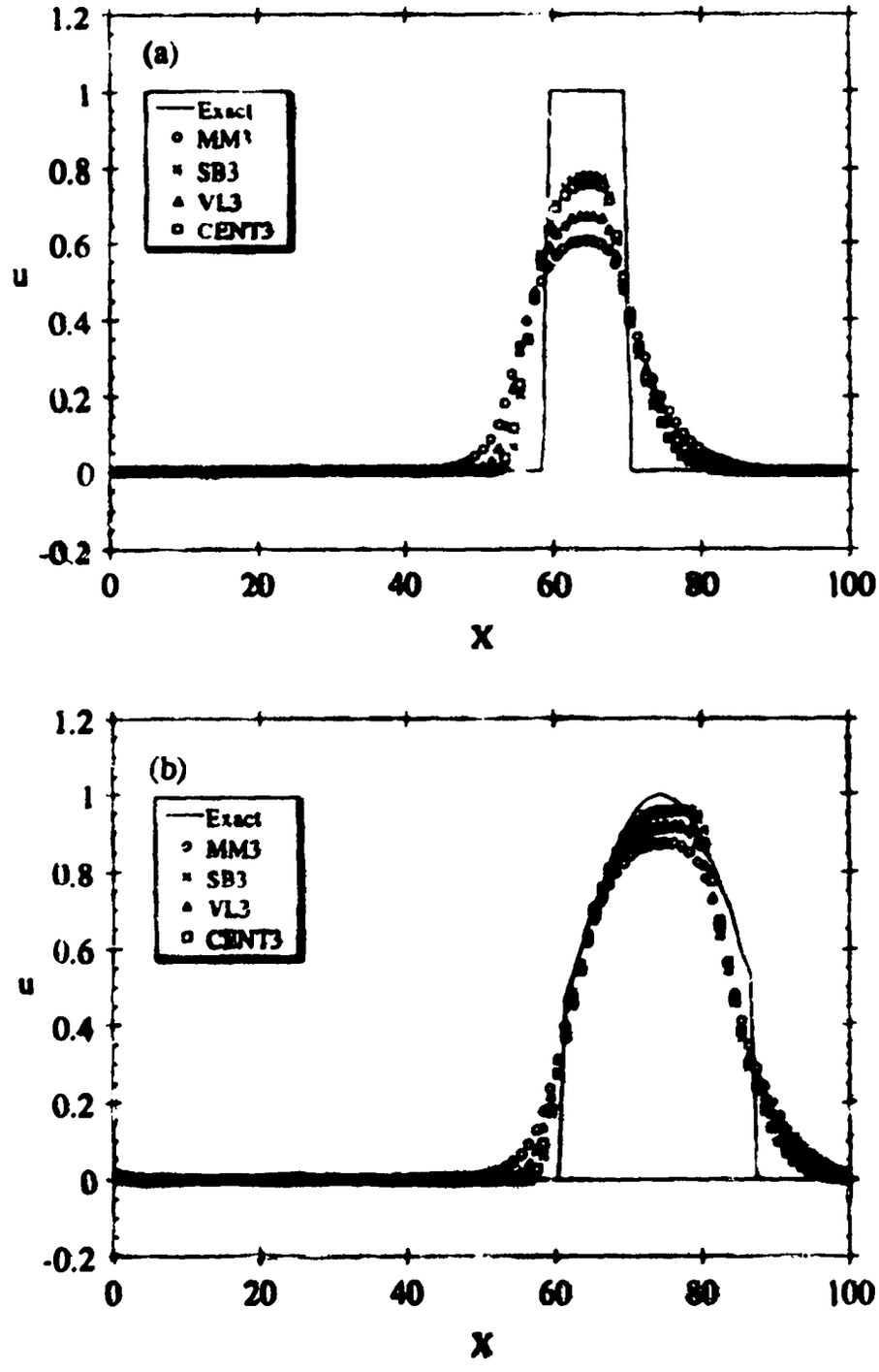


Figure 8.22: The scalar square and $\sin^2 x$ wave solutions using several three argument TVD limiters.

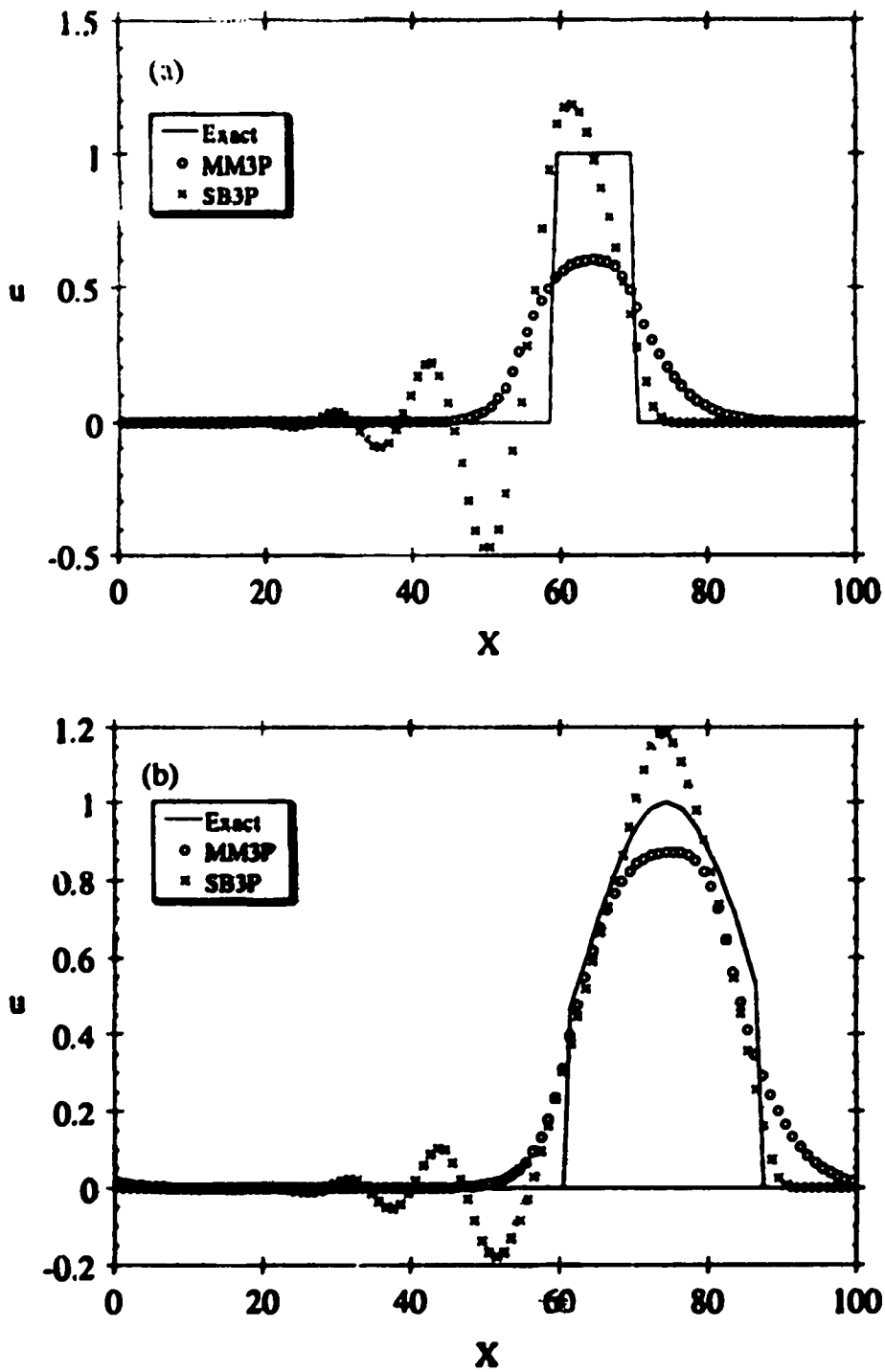


Figure 8.23: The scalar square and $\sin^2 x$ wave solutions using several three argument "prime" limiters. Note the decidedly non-TVD behavior of the SB3P limiter.

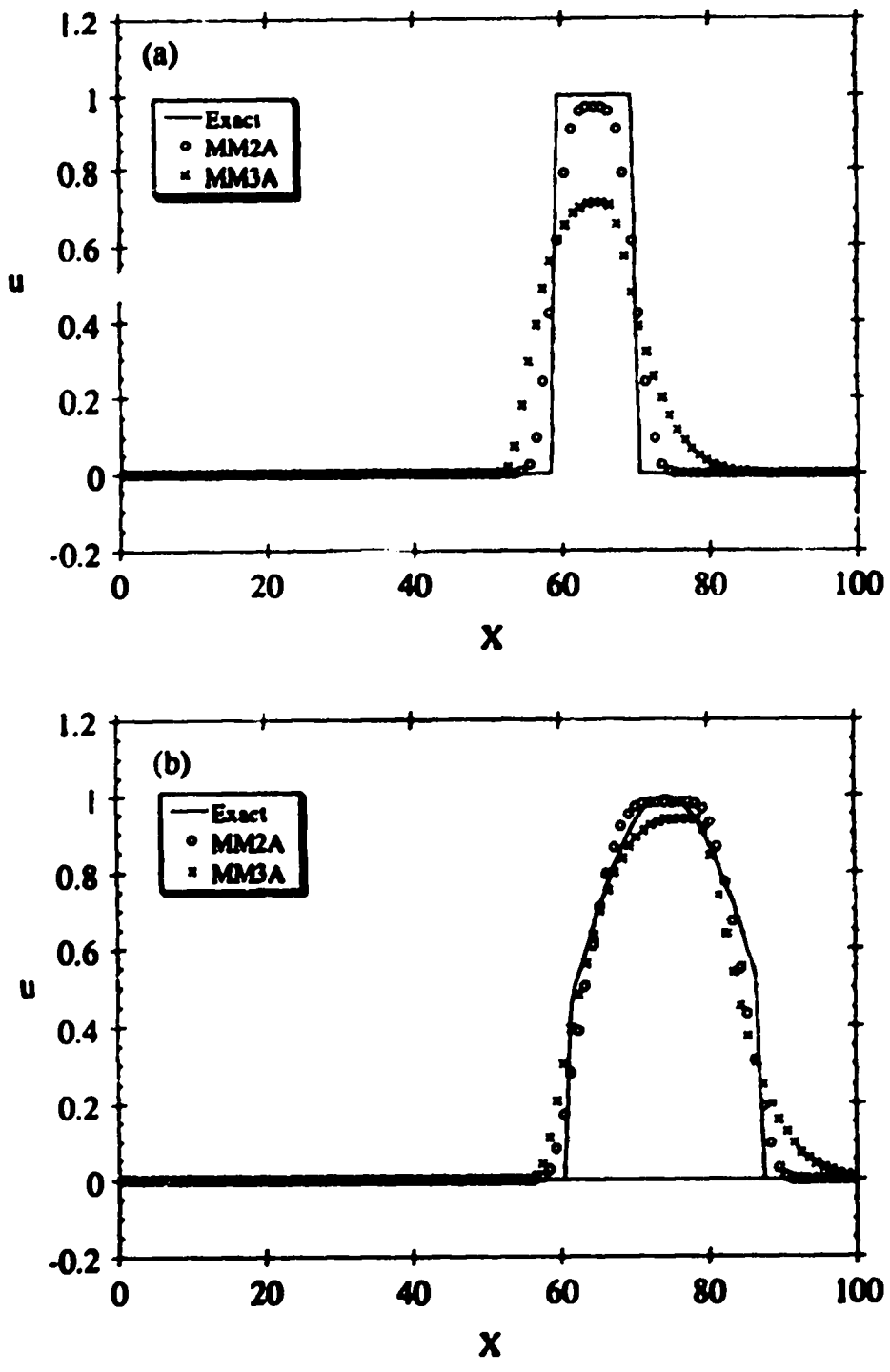


Figure 8.24: The scalar square and $\sin^2 x$ wave solutions using artificial compression. It is notable that the solution with the two argument limiters (MM2A) compresses the $\sin^2 x$ profile in a similar manner to the SB2 limiter.

iter performs quite well, improving the resolution of the basic two argument limiter-based solution at the cost of not being TVD. The three argument TVB limiter does not fair as well. This can be attributed to the "local" nature of the resulting scheme, which looks too much like the Lax Wendroff scheme. In Fig. 8.25, the MM3TVB is virtually identical to the corresponding Lax-Wendroff solution. To combat this problem, two other forms of the three argument limiter are introduced, the MM3TVB'

$$Q^{TVB}\left(r^-,1,r^+\right) = \max\left[0,\min\left(r^- + m,1 + m,r^+ + m,\frac{1}{2}\left(r^- + r^+\right)\right)\right] . \quad (8.51a)$$

and MM3TVB"

$$Q^{TVB}\left(r^-,1,r^+\right) = \max\left[0,\min\left(r^- + m,1 + m,r^+ + m,\frac{1}{2}\left(1 + r^+\right),\frac{1}{2}\left(r^- + 1\right)\right)\right] . \quad (8.51b)$$

As Fig. 8.27 shows, the results are improved. The tabular data also reveal this.

Figure 8.28 shows the results obtained with S-limiters. For the two argument case, the results are not significantly different than those obtained with standard TVD two argument limiters. The S-limiters have a slight advantage in terms of the quality of results with slightly lower numerical diffusion. As revealed by looking at the numerical data, the three argument case is improved greatly by the use of the S-limiters when compared with the corresponding TVD limiter case. This is most likely due to some reduction in the clipping of smooth extrema in the solution.

Van Albada's limiter is used to represent the solution by a generalized average limiter ($n = 2$). I have already seen the van Leer or $n = 1$ limiter at work. The results in Fig. 8.29 do not use bias in the schemes. The results are quite comparable with other two or three argument TVD type schemes. In fact, the solutions are quite similar to those obtained with the VL2 or VL3 limiters. By adding bias to the limiter, the resolution can be improved in a qualitative sense. In a quantitative sense, the results are worse. One interesting remark is that the three argument limiters in general seem to be more sensitive (as seen in this case or the TVB limiters)

8.4.2 Burgers' Equation

This section of the chapter centers on the order of accuracy obtained with methods in conjunction with limiters and their subsequent solutions. To accomplish this, a standard test problem using Burgers' equation is used. The problem consists of an initial condition of $\sin(x)$, $x \in [0, 2\pi]$. At $t = 0.2$, the solution is smooth, and at $t = 1.0$, a shock has formed in the solution. It is at these times that the accuracy of the solution is assessed. The problem is solved with 10 grid cells followed by 1000 grid cells. The solution is obtained with a Godunov numerical fluxes as described in [158].

The results for this test problem are given in Tables 8.6 and 8.7. In general, the

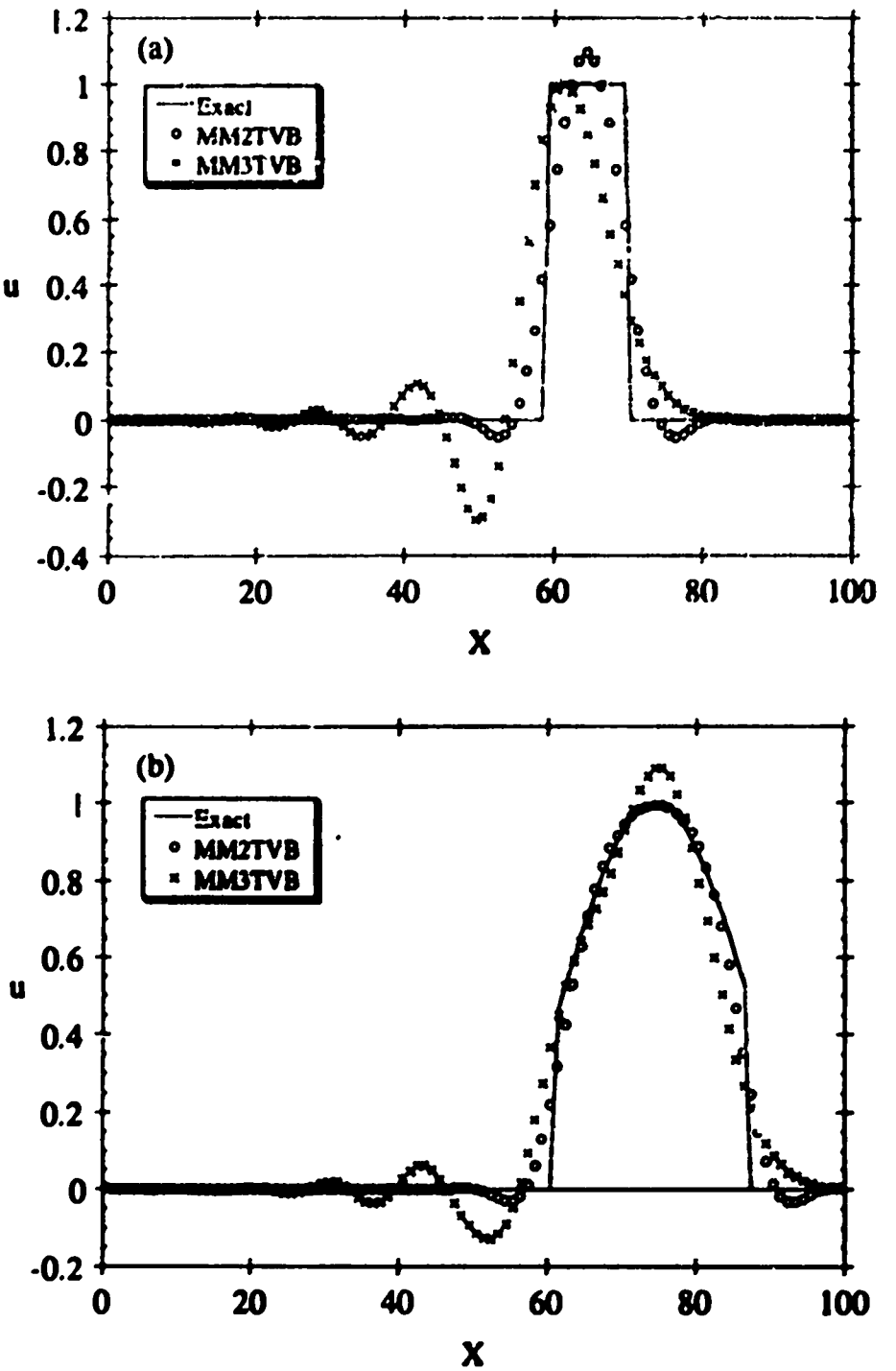


Figure 8.25: The scalar square and $\sin^2 x$ wave solutions using TVB limiters. The three argument TVB limiter produces a results nearly identical to the Lax-Wendroff method.

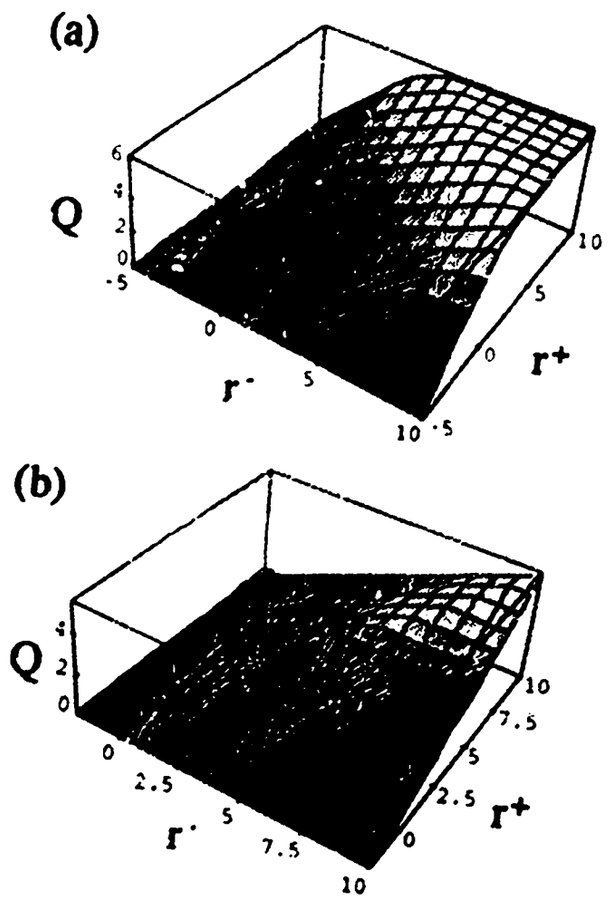


Figure 8.26: The modified three argument TVB limiter is shown here for $M\Delta x = 5$. MM3TVB' is shown in Fig. 8.26a. MM3TVB" is shown in Fig. 8.26b.

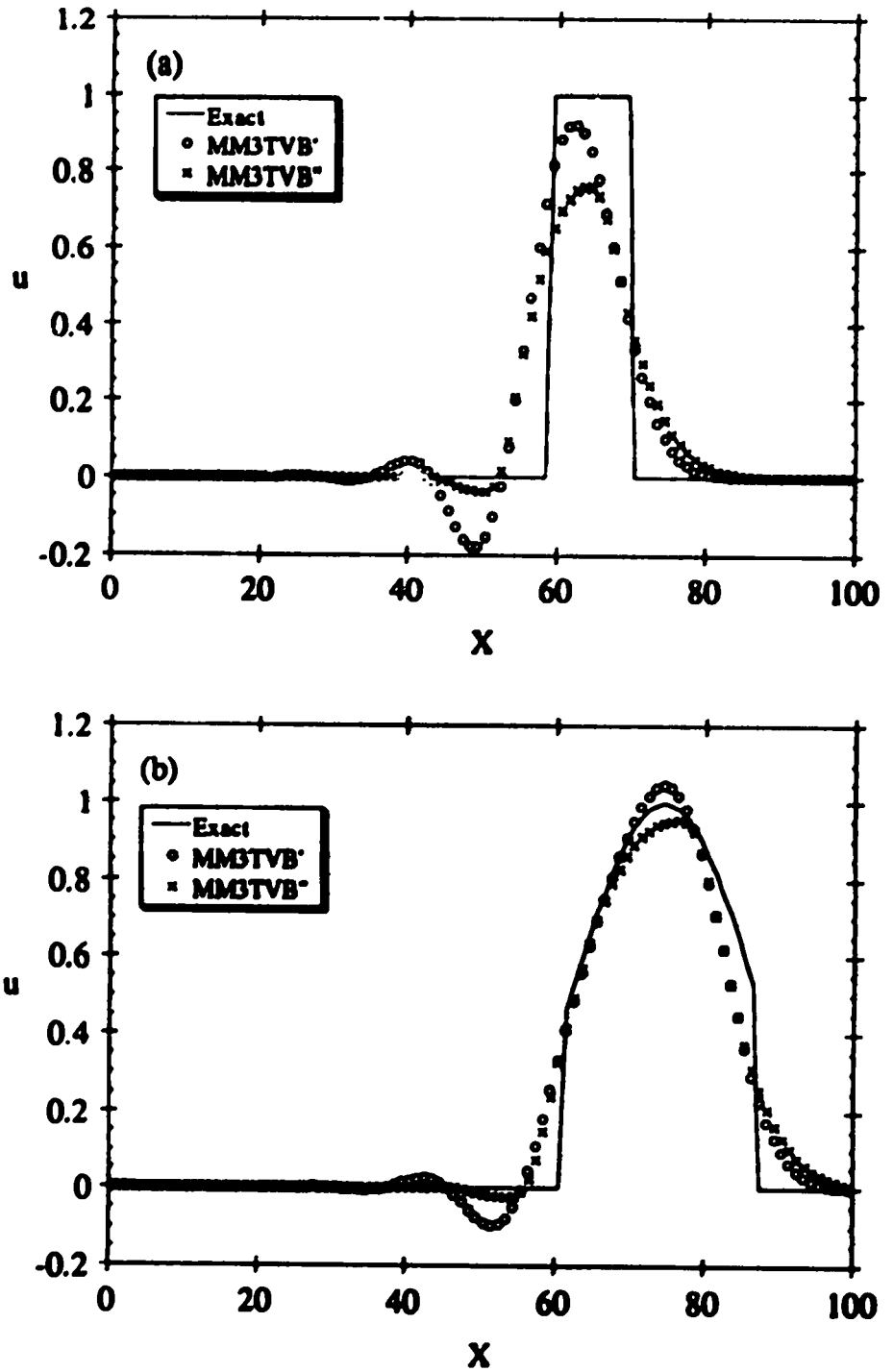


Figure 8.27: The scalar square and $\sin^2 x$ wave solutions using modified three argument TVB limiters. These improve the performance of the three argument TVB limiters.

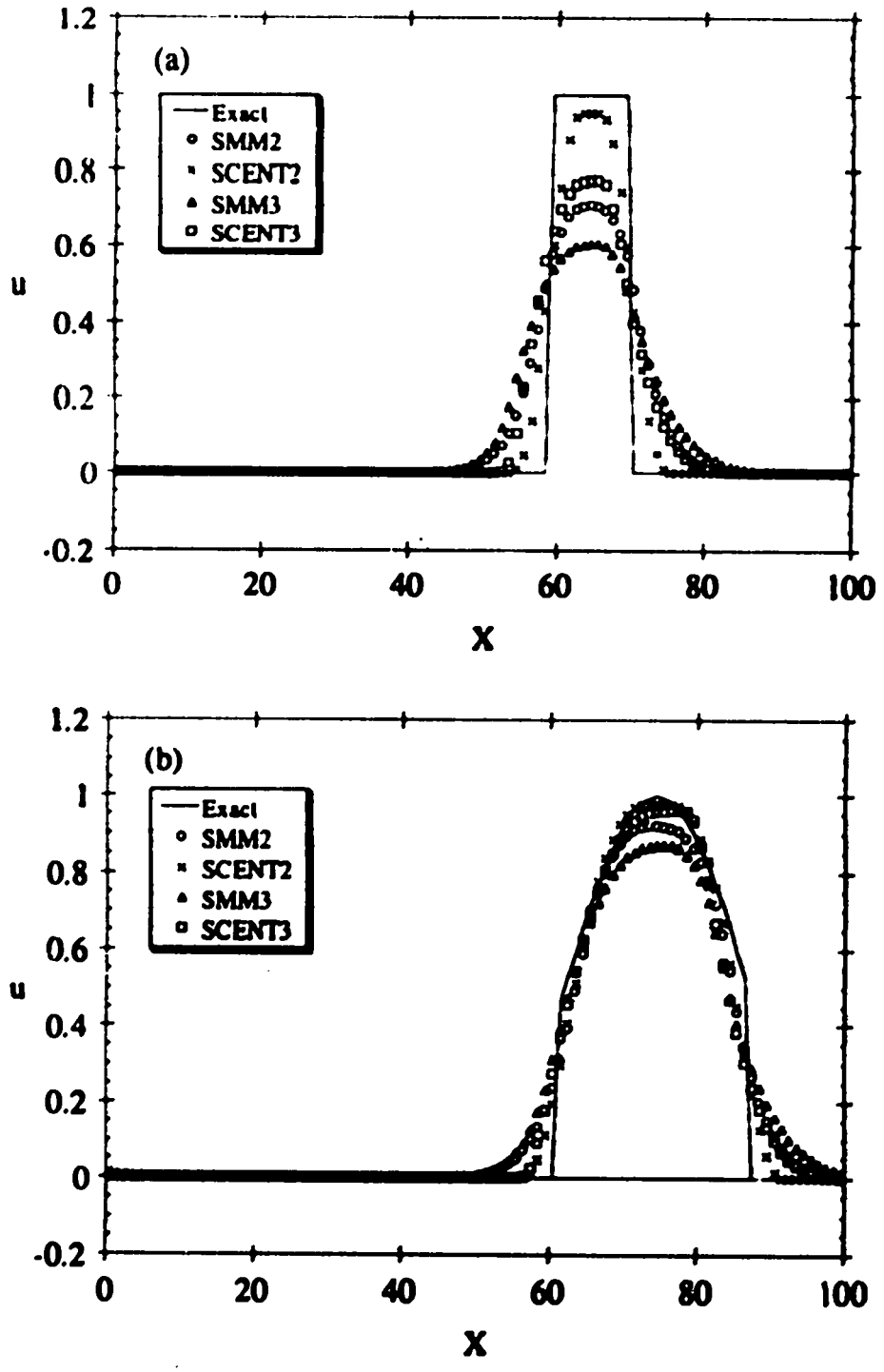


Figure 8.28: The scalar square and $\sin^2 x$ wave solutions using two and three argument S-limiters.

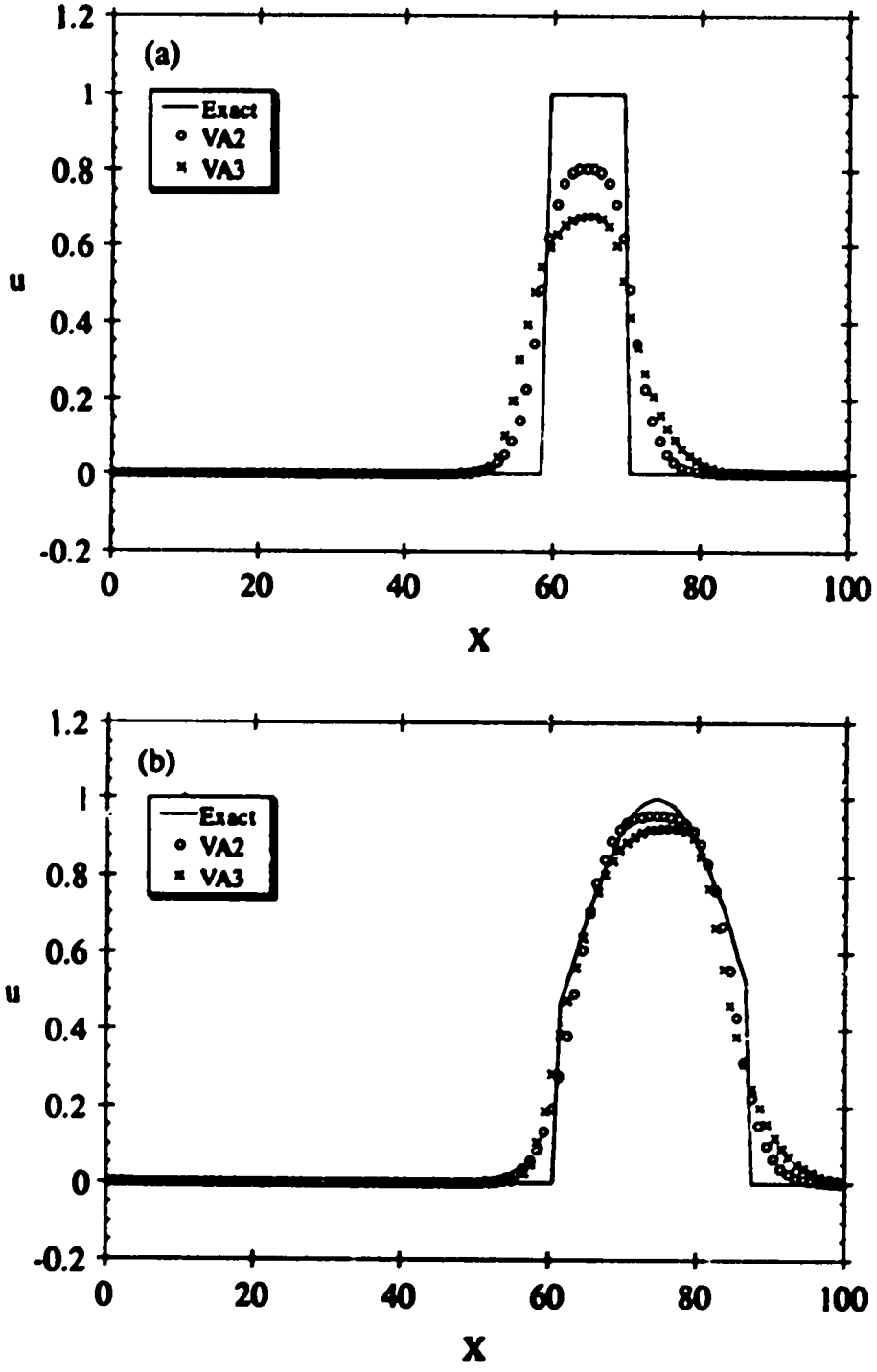


Figure 8.29: The scalar square and $\sin^2 x$ wave solutions using the generalized average limiters with $n = 2$.

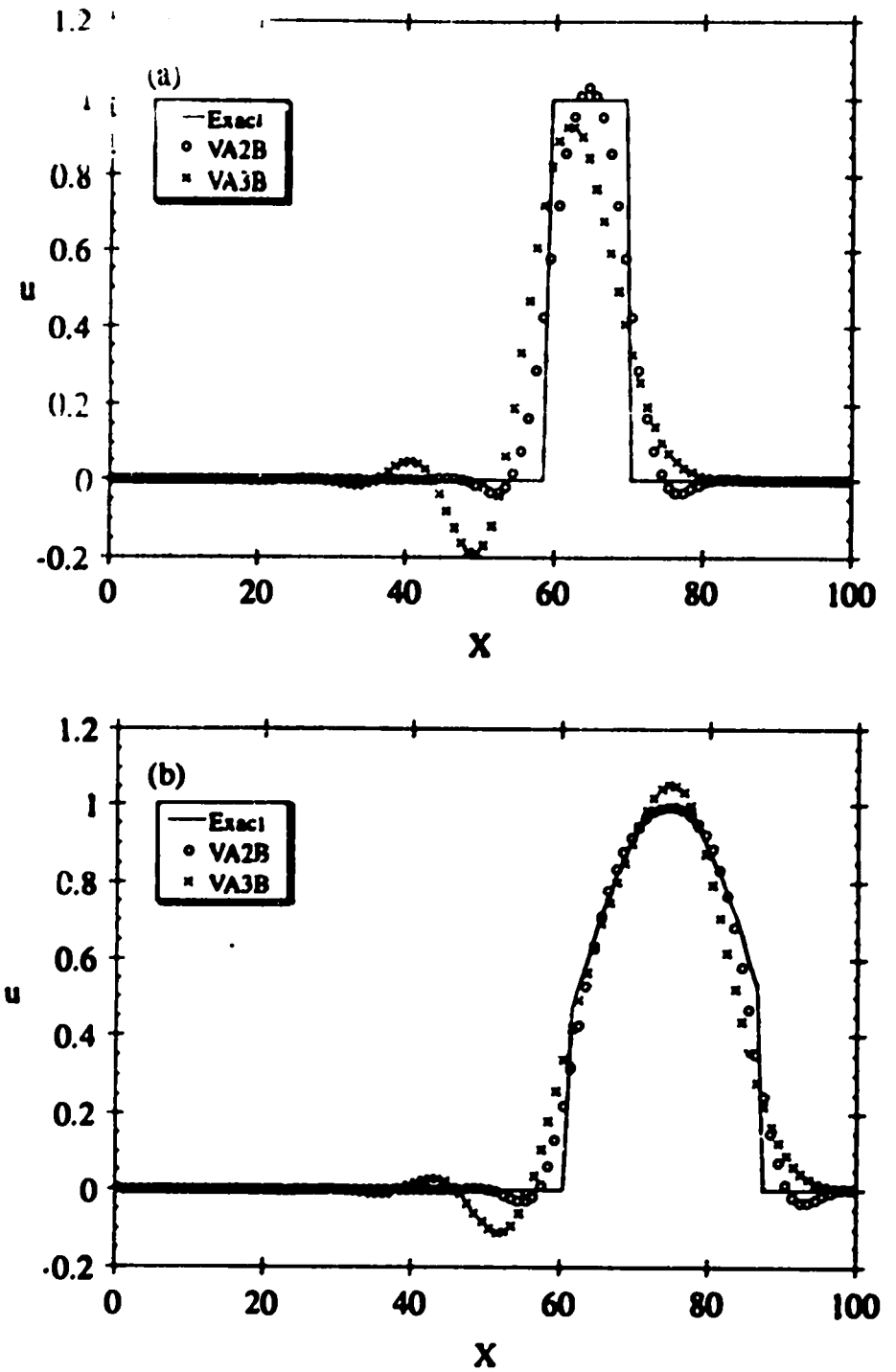


Figure 8.30: The scalar square and $\sin^2 x$ wave solutions using the generalized average limiters with $n = 2$ with a bias added as suggested in [198].

Table 8.3: L_1 error norms with minimum and maximum values for the square wave problem.

Limiter	Minimum	Maximum	L_1 error
MM2	0.0000	0.7108	7.41×10^{-2}
VL2	0.0000	0.8784	4.59×10^{-2}
CENT2	0.0000	0.9508	3.65×10^{-2}
SB2	0.0000	0.9927	1.79×10^{-2}
MM3	0.0000	0.6037	9.41×10^{-2}
MM3P	0.0000	0.6005	9.47×10^{-2}
SB3	0.0000	0.7819	6.36×10^{-2}
SB3P	-0.1690	1.1875	9.71×10^{-2}
VL3	0.0000	0.6760	8.20×10^{-2}
CENT3	0.0000	0.7632	6.60×10^{-2}
MM2A	0.0000	0.9668	3.14×10^{-2}
MM3A	0.0000	0.7174	7.55×10^{-2}
MM2TVB	-0.0514	1.0901	4.00×10^{-2}
MM3TVB	-0.0392	0.7616	7.77×10^{-2}
SMM2	0.0000	0.7108	7.41×10^{-2}
SCENT2	0.0000	0.9516	3.65×10^{-2}
SMM3	0.0000	0.6059	9.39×10^{-2}
SCENT3	0.0000	0.7758	6.52×10^{-2}
VA2	0.0000	0.8035	5.63×10^{-2}
VA3	0.0000	0.6801	7.95×10^{-2}
VA2B	-0.0314	1.0313	4.04×10^{-2}
VA3B	-0.1885	0.9275	7.78×10^{-2}

Table 8.4: L_1 error norms with minimum and maximum values for the $\sin^2 x$ wave problem.

Limiter	Minimum	Maximum	L_1 error
MM2	0.0000	0.9197	3.74×10^{-2}
VL2	0.0000	0.9668	2.26×10^{-2}
CENT2	0.0000	0.9794	1.94×10^{-2}
SB2	0.0000	0.9893	2.43×10^{-2}
MM3	0.0000	0.8717	5.20×10^{-2}
MM3P	0.0000	0.8708	5.24×10^{-2}
SB3	0.0000	0.9552	2.98×10^{-2}
SB3P	-0.1801	1.1847	5.63×10^{-2}
VL3	0.0000	0.9162	4.06×10^{-2}
CENT3	0.0000	0.9571	3.00×10^{-2}
MM2A	0.0000	0.9835	2.10×10^{-2}
MM3A	0.0000	0.9385	3.53×10^{-2}
MM2TVB	-0.0321	0.9943	2.08×10^{-2}
MM3TVB	-0.0266	0.9538	3.95×10^{-2}
SMM2	0.0000	0.9195	3.74×10^{-2}
SCENT2	0.0000	0.9791	1.95×10^{-2}
SMM3	0.0000	0.8726	5.20×10^{-2}
SCENT3	0.0000	0.9606	3.00×10^{-2}
VA2	0.0000	0.9524	2.59×10^{-2}
VA3	0.0000	0.9217	3.56×10^{-2}
VA2B	-0.0319	0.9944	2.02×10^{-2}
VA3B	-0.1086	1.0564	4.37×10^{-2}

Table 8.5: Numerical viscosity and total variation for both scalar wave equation problems.

Limiter	$\sum \tau$ square	TV square	$\sum \tau \sin^2 x$	TV $\sin^2 x$
MM2	40.67	1.42	30.61	1.84
VL2	17.65	1.76	7.91	1.93
CENT2	10.74	1.90	3.58	1.96
SB2	3.00	1.99	-8.49	1.98
MM3	60.59	1.21	53.15	1.74
MM3P	61.19	1.20	53.62	1.74
SB3	30.57	1.56	17.52	1.91
SB3P	-26.63	4.052	-23.78	3.11
VL3	47.09	1.35	35.02	1.83
CENT3	31.97	1.53	17.91	1.91
MM2A	8.19	1.94	-1.38	1.97
MM3A	40.73	1.43	29.39	1.88
MM2TVB	7.90	2.41	3.36	2.12
MM3TVB	39.71	1.61	29.39	1.96
SMM2	40.47	1.42	30.55	1.84
SCENT2	10.63	1.90	3.53	1.96
SMM3	60.09	1.21	52.94	1.75
SCENT3	30.83	1.55	17.53	1.92
VA2	25.70	1.61	12.72	1.90
VA3	44.75	1.36	39.82	1.84
VA2B	9.11	2.20	3.37	2.13
VA3B	12.11	2.37	4.38	2.42

Table 8.6: Order of convergence in several error norms for Burgers' equation at $t = 0.2$ when the solution is smooth.

Limiter	L_1	L_2	L_∞
MM2	2.12	2.15	1.84
VL2	2.15	2.17	1.84
CENT2	2.16	2.17	1.95
SB2	2.18	2.17	1.84
MM3	2.08	1.86	1.32
MM3P	2.08	1.87	1.32
SB3	2.15	1.85	1.31
SB3P	1.91	1.63	1.08
VL3	2.12	1.85	1.31
CENT3	2.13	1.86	1.32
MM2A	2.14	2.16	1.83
MM3A	2.12	1.84	1.31
MM2TVB	1.73	1.73	1.63
MM3TVB	2.04	1.82	1.28
SMM2	2.12	2.14	1.85
SCENT2	2.16	2.16	1.83
SMM3	2.08	1.84	1.27
SCENT3	2.08	1.81	1.26
VA2	2.16	2.18	1.87
VA3	2.13	1.86	1.31
VA2B	1.73	1.74	1.64
VA3B	2.02	1.80	1.25

Table 8.7. Order of convergence in several error norms for Burgers' equation at $t = 0.2$ when the solution has a shock in it.

Limiter	L_1	L_2	L_∞
MM2	1.47	1.12	0.70
VL2	1.51	1.10	0.61
CENT2	1.52	1.10	0.61
SB2	1.51	1.01	0.49
MM3	1.57	1.18	0.74
MM3P	1.57	1.18	0.74
SB3	1.68	1.14	0.60
SB3P	1.28	0.79	0.25
VL3	1.65	1.19	0.69
CENT3	1.53	1.00	0.47
MM2A	1.49	1.08	0.58
MM3A	1.60	1.10	0.54
MM2TVB	1.19	0.83	0.36
MM3TVB	1.52	1.05	0.51
SMM2	1.51	1.14	0.70
SCENT2	1.60	1.16	0.63
SMM3	1.51	1.15	0.72
SCENT3	1.52	0.98	0.44
VA2	1.54	1.12	0.65
VA3	1.65	1.13	0.60
VA2B	1.15	0.77	0.31
VA3B	1.51	1.01	0.48

order of convergence for the solutions is better for the two argument limiter than the three argument limiters. The three argument limiters also experience a much greater difference in convergence from one norm to a higher norm. The non-TVD and FCT limiters seem to suffer from worse convergence characteristics than the other schemes. Additionally, the schemes using some constant (TVB or VA2B and VA3B) in the limiter show poor convergence. These schemes do perform far better when the mesh is coarse, and these limiters seem to produce excellent results in relation to other limiters for those cases. After a shock has formed, the two argument limiters show a greater degradation in convergence. Again, this is especially true with the non-TVD limiters. The stated convergence of the three argument limiters when a shock has formed is somewhat a function of the exceedingly poor results found on the coarsest grid. In the same vein, the poor convergence of the TVB and the biased van Albada limiters is somewhat a result of the excellent results obtained on the coarsest grid.

8.5 Concluding Remarks

In this chapter a number of limiters have been reviewed and their properties examined. In addition, several limiters have been introduced or reformulated and analyzed within a common framework The impact of limiters on high-resolution numerical solutions has also been demonstrated. The importance of limiters on the solution of the equations is undeniable. The quality of solutions is directly traceable to the limiters because they are the heart of the numerical schemes.

More study of limiters is warranted in light of these results. As discussed earlier, limiters can impact steady-state solution convergence. Some study of this phenomena is needed. Additionally, both TVB and generalized average limiters should studied in order to give more systematic manner to choose the constants used with the limiters.

The following chapter explores the effect of the constraints placed on the polynomial interpolation employed by high-order Godunov schemes.

Chapter 9.

Cell-Averages or Point-Values? On Reconstruction Methods

> We have found a strange footprint on the shores of the unknown. We have devised profound theories, one after another, to account for its origin. At last we have succeeded in reconstructing the creature that made the footprint. And lo! it is our own. *Sir Arthur Stanely Eddington*

9.1 Introduction

One of the primary manners of constructing modern high-resolution upwind schemes is the use of the HOG philosophy. This method has several key points in its favor: the use of conservation form, the ease of use with systems of equations, the use of a quality underlying physical model, and reduction of finite differences to polynomial interpolation. It is this final point on which I concentrate my efforts.

The polynomial reconstruction determines the order of accuracy the scheme can attain. It also interacts strongly with the underlying physical model mentioned in the previous paragraph. This underlying model is typically a Riemann solver of some variety [30]. In HOG methods, the polynomials used are constructed piecewise so that each control volume has one polynomial per variable in it. At the boundaries of the control volume, the polynomial distributions are not required to be continuous and a discontinuity typically results. The Riemann solver acts as a sort of "referee" determining what the correct numerical flux should be at that cell boundary. I return to the general description of HOG methods in the following section.

These methods grew out of the work of Godunov [56, 57] whose ingenious method embodied the essence of upwind differencing as given by Courant Issacson, and Rees [54, 31]. The work of Godunov was important in two regards because of his use of a Riemann solver within the difference scheme and his theorem regarding difference schemes.

In the 1970s, a number of researchers made great strides in using nonlinear schemes in attaining monotone schemes of higher order accuracy. Notable among these works is that of Boris and Book [59] on the flux-corrected transport method and Harten's artificial compression method [183]. The work of van Leer was connected more closely to that of Godunov and in a series of papers, HOG methods were defined [119, 120, 60]. The key to this definition was the definition of monotone advection using higher order polynomial descriptions of the numerical flux.

Van Leer's work on HOG methods was extended in a number of efforts in the 1980s. The PPM [122, 27] is notable because of its continued preeminence in the field [129, 89]. This method was originally conceived in Lagrangian coordinates coupled with an Eulerian remap, but is equally at home in purely Eulerian coordinates [123]. A significant advance in HOG methods was made with ENO methods [61, 65, 66]. These methods extended the HOG method to an arbitrary high order of accuracy.

Perhaps of equal importance to the development of HOG method has been the advent of TVD [130, 61] and TVB [169] methods. The criteria defining methods to be either TVD or TVB apply to HOG methods. The class of TVD (and consequently TVB) methods is quite large and it is usually not difficult to show a direct correspondence between these methods and the HOG methodology. This idea is key to the analysis that follows.

This chapter is divided into five sections. The second section gives a basic introduction to HOG methods. The following section describes basic cell-average and point value algorithms considered here. This is followed by a presentation and discussion of the performance of the methods. The fifth section has conclusions and closing remarks.

9.2 High-Order Godunov Methods

As noted above, the HOG methods use a nonlinear piecewise polynomial interpolation to define their numerical fluxes in conjunction with a Riemann solver.

The schematic representation of a second-order method is shown in Fig. 9.1. As can be seen by comparing this with Fig. 4.5, the only difference between them is in the reconstruction step, which in turn impacts the solution in the small.

A numerical flux is determined by the two states meeting at any given cell edge and the Riemann solver. In Godunov's method, the cell-edge values are equal to the cell-average (or cell center) values. Thus, a zeroth order polynomial describes the variables distribution in any given cell. An integration of the reconstructive polynomial over the cell trivially recovers the cell-average.

At this point in the exposition, it is helpful to concretely state what is meant by a cell-average or point-value-based interpolation.

Definition 7 (cell-average reconstruction) *A piecewise polynomial reconstruction is cell-average based if the average of the reconstruction over the cell is equal to the cell-average.*

Definition 8 (point-value reconstruction) *Point-value interpolation is more loosely defined. The piecewise polynomial reconstruction is a polynomial of some accuracy interpolating the data within a given cell.*

The polynomial in the point-value reconstruction can obey any number of possible constraints based on various derivatives, moments, and quality factors related to

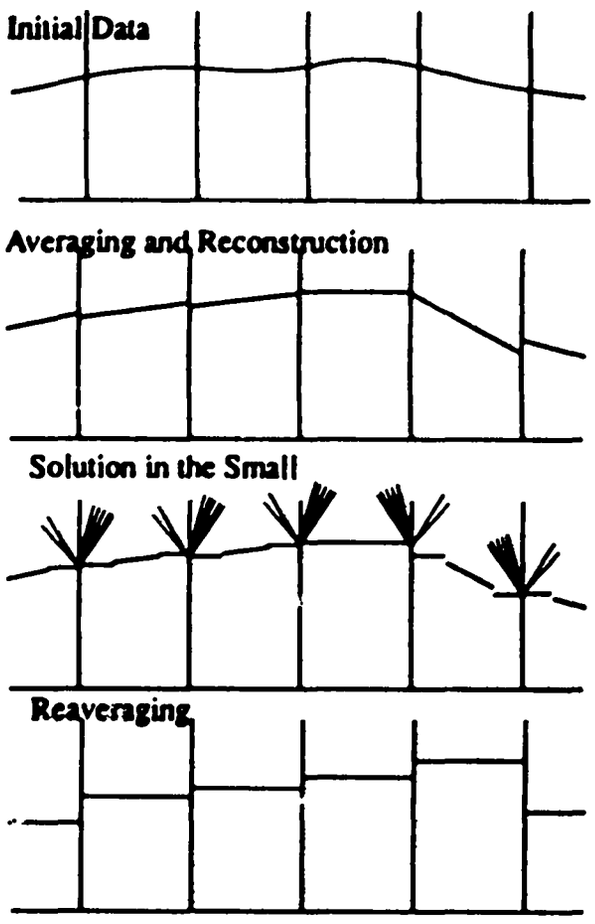


Figure 9.1: The steps of Godunov's methods are shown for a higher order polynomial reconstruction. The solution in the small takes place with data that has been time centered over the domain of dependence of the local characteristics.

the data. In a sense, the cell-average reconstruction is a subset of the point-value reconstruction based on the restriction to an integral constraint based on a cell-average.

At this point, a deeper meaning should be gleaned from the above presentation of Godunov's method. Godunov's method has as its basis the idea of cell-averages. The cell-averages of the dependent variables are conserved by the scheme. In the above algorithm, the cell-averages represent the quantities used in the method derivation. The use of cell-averages fits nicely into the theory of weak solutions given in Chapter 2.

The question to ponder is whether it is necessary for the cell-averages to be used exclusively in the difference schemes. The conservation form of the finite difference scheme ensures that the cell-averages are conserved. The key question regards the accuracy and efficiency of the approximation. At a more philosophical level, the generality of the design principle comes to play. Because the point-value philosophy is more general it lends itself to extension in multiple dimensions and other types of problems with greater ease than the more restrictive cell-average reconstruction.

The formulation of Godunov's method implies the use of some cell-average interpolation. The use of the divergence theorem to transform the integrals to forms more amenable to numerical treatment changes the situation somewhat. It is necessary to compute the flux functions in order to compute changes in the cell-averages. The conservation is not effected by this change regardless of the method used to compute the fluxes (as long as $f_{j+\frac{1}{2}} = f_{j+\frac{1}{2}}$ irregardless of what cell is being computed). The upwind principles embodied by Riemann solvers and appropriate monotonicity constraints on the reconstruction ensure that the fluxes are of a quality nature.

Point-values of the function being advected should be reasonable representations of the function in any given control volume and by the mean value theorem should be fairly close to the cell-averages. As noted in [199], the cell averages and point values differ by $\mathcal{O}(\Delta x^2)$. These values should certainly be acceptable for the computation of fluxes, because the form of the difference equations conserves the cell-averages. Most classical difference are based on point-value interpolation (or can be thought of in this way).

The cell-average basis makes good theoretical and logical sense. Given a finite volume discretization and taking into account a Gibbs-type error would imply that you could only know the cell-averages. The point of importance is how to construct a piecewise reconstruction for the purposes of computing fluxes.

9.3 Description of Polynomial Reconstructions

As noted in the previous section, I examine two approaches to reconstruction in HOC methods. The cell-average formulation is more theoretically pleasing, but the point-value formulation is simpler and seems more natural at first glance.

9.3.1 Cell-Average Reconstruction

This section of the chapter concerns the construction of piecewise polynomials of the cell-average type.

The canonical cell-average reconstruction is used in Godunov's method, i.e.,

$$P(x) = u_j \ , \ x \in \left[x_{j+\frac{1}{2}}, x_{j+\frac{1}{2}}\right] \ . \tag{9.1}$$

This method has first-order accuracy and trivially has the cell-average reconstruction property.

A second-order method widely used for HOG type algorithms [123, 179] is defined by the reconstructive polynomial

$$P(x) = u_j + \widetilde{\Delta_j u_j} \frac{(x - x_j)}{\Delta_j x} \ , \ x \in \left[x_{j+\frac{1}{2}}, x_{j+\frac{1}{2}}\right] \ . \tag{9.2}$$

The slope, $\widetilde{\Delta_j u}/\Delta_j u$, is a limited estimate of du/dx at the cell center, x_j. The limiters used were discussed in Chapter 8. Integration over the cell confirms that this reconstruction has the cell-average property. This scheme is compared with a point-value type of reconstruction in Section 9.4.

The third form of cell-average reconstruction is the MUSCL reconstruction [120, 147, 45]. This form is particularly useful because it has a parametric form and thus is actually a family of schemes. The polynomial is based on Legendre polynomials, and thus has the desired cell-average reconstruction property. The basic form of the scheme's reconstruction is

$$\begin{aligned} P(x) = u_j &+ \frac{1}{2}\left(\tilde{s}_{j-\frac{1}{2}} + \tilde{s}_{j+\frac{1}{2}}\right)(x - x_j) \\ &+ \frac{3\kappa}{2}\left(\tilde{s}_{j+\frac{1}{2}} - \tilde{s}_{j-\frac{1}{2}}\right)\left[\frac{(x - x_j)^2}{\Delta x} - \frac{\Delta x^2}{12}\right] \ , \ x \in \left[x_{j+\frac{1}{2}}, x_{j+\frac{1}{2}}\right] \ . \end{aligned} \tag{9.3}$$

Here $\tilde{s}_{j-\frac{1}{2}} = Q(l, r)s_{j+\frac{1}{2}}$ where $Q(l, r)$ is a limiter and $s_{j-\frac{1}{2}} = \Delta_{j-\frac{1}{2}}u/\Delta_{j-\frac{1}{2}}x$. Table 9.1 gives the types of schemes that arise for different values of κ. Care must be taken in the use of limiters with this scheme, as was discussed previously (Section 8).

One problem with this scheme is the definition of the stencil used for the limiters. If the stencil is not chosen correctly, the scheme, although stable, is not TVD, and thus be oscillatory. In general, upwind biased limiters used with this scheme do not produce TVD results because the upwind biased gradients used in defining the reconstruction apply their information throughout the cell, thus violating the assumptions made with an upwind biased stencil. This problem can be cured through centering the stencil in some manner. One option is to center the limiters, but this has a detrimental impact on the scheme's resolution.

Before moving onto point-value based reconstructions, some comments must be

Table 9.1: The type of scheme produced for various values of κ with the MUSCL reconstruction.

κ	Scheme
-1	one-sided, second-order
0	upwind, second-order
1/3	upwind, third-order
1	centered, second-order

made concerning ENO type schemes. The powerful PPM method is based on cell-average reconstruction. This scheme uses a quadratic cell-average reconstruction. Another concept used with this scheme is a primitive function that is used to define values of u at the cell interfaces. The primitive function of u is defined by

$$U(x) = \int_{-\infty}^{x} u(x)\,\Delta x\,. \tag{9.1}$$

This concept is put to greater use in the derivation ENO schemes [64]. The actual reconstruction takes place with the primitive function. This reconstruction is then differentiated to give the reconstruction to $u(x)$. By inspection, this scheme has the cell-averaged reconstruction property. One important caveat is that this does not generalize to multiple dimensions except through dimensional splitting. This is due to the lack of a generalization of the primitive function concept to multidimensional cases.

To test the cell-average reconstruction I used two test problems: one with a smooth nearly discontinuous form, and a second with a smooth local extrema. The first problem was used to test the PPM [122, 27] method, and has the functional form

$$f(x) = \tanh(x)\ ,$$

the second problem is a Gaussian distribution with a standard deviation $\Delta x = 3$

$$f(x) = \exp\left[-\left(x^2\right)/2\Delta x\right]\ .$$

Both are plotted over the range $x \in [-10, 8]$.

The results for these functions with Godunov's method are shown in Fig. 9.2. The large jumps result in a large amount of diffusion in the solution as given by the theory shown in Chapter 8. By going to a second-order algorithm, the results improve. Figure 9.3 shows the basic second-order HOG algorithm with the minmod limiter

The diffusion has been decreased because the jumps have diminished in magnitude. By using the central limiter these results improve, and by using the superbee limiter the results improve again. This is shown in Figs. 9.4 and 9.5. With the Gaussian distribution, the superbee has overcompressed one location, which is typically the beginning of forming a false discontinuity. The use of cell-averages is diffusive, (in fact TVD see [64]) and results in the immediate clipping of an extrema in the solution.

Figure 9.6 shows the reconstruction using the MUSCL interpolant with $\kappa = 1/3$. The use of three argument limiters makes this a TVD scheme, but as noted in Chapter 8 the three argument limiters are more diffusive than the two argument limiters. The $\tanh(x)$ grid is too coarse to capture the discontinuity with these limiters.

The methods for reconstruction given above aree contrasted with the methods discussed in the following section for form and complexity.

9.3.2 Point-Value Reconstruction

In this section, I introduce the general concept in point-value based reconstruction and compare some specific examples with the cell-average formulation in Section 9.4.

If, for instance, the cell-averages are not used to derive the fluxes, the scheme still maintains its conservation. The canonical example of this is the Lax-Wendroff method. This method is conservative, but its HOG analog described in Chapter 6 does not use a reconstruction, which is of a cell-average variety.

The integral average of the Lax-Wendroff polynomial over a cell $x \in \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$ yields

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} P_j(x)\,dx = u_j + \frac{\Delta x}{8}\left(s_{j+\frac{1}{2}} - s_{j-\frac{1}{2}}\right), \tag{9.5}$$

which does not equal u_j unless $s_{j-\frac{1}{2}} = s_{j+\frac{1}{2}}$.

With the inclusion of slope limiters, this scheme becomes the symmetric HOG method (see Chapter 6). These limiters can either be upwind biased or centered in their support (see Chapter 8) . These schemes are defined by changing $s_{j\pm\frac{1}{2}} \rightarrow \tilde{s}_{j\pm\frac{1}{2}}$ in (3.12a). Here $\tilde{s}_{j\pm\frac{1}{2}}$ are defined with appropriate limiters [132, 134].

In Chapter 6, the scheme above was extended to include a quadratic interpolation based on the same available data (one degree of freedom is not used in the above schemes). Although not stated in Chapter 6, this scheme is the analog to the MUSCL reconstruction using Taylor rather than Legendre polynomials. This scheme is described by the reconstruction

$$P(x) = u_j + \frac{1}{2}\left(\tilde{s}_{j-\frac{1}{2}} + \tilde{s}_{j+\frac{1}{2}}\right)(x - x_j) + \kappa\left(\tilde{s}_{j+\frac{1}{2}} - \tilde{s}_{j-\frac{1}{2}}\right)\frac{(x - x_j)^2}{\Delta x}, \quad x \in \left[x_{j+\frac{1}{2}}, x_{j+\frac{1}{2}}\right]. \tag{9.6}$$

The lower operation count in the above equation is evident by comparing the two forms. The family of schemes produced for differing values of κ is described by Table 9.2. In the following section, the limiters used with these schemes are discussed.

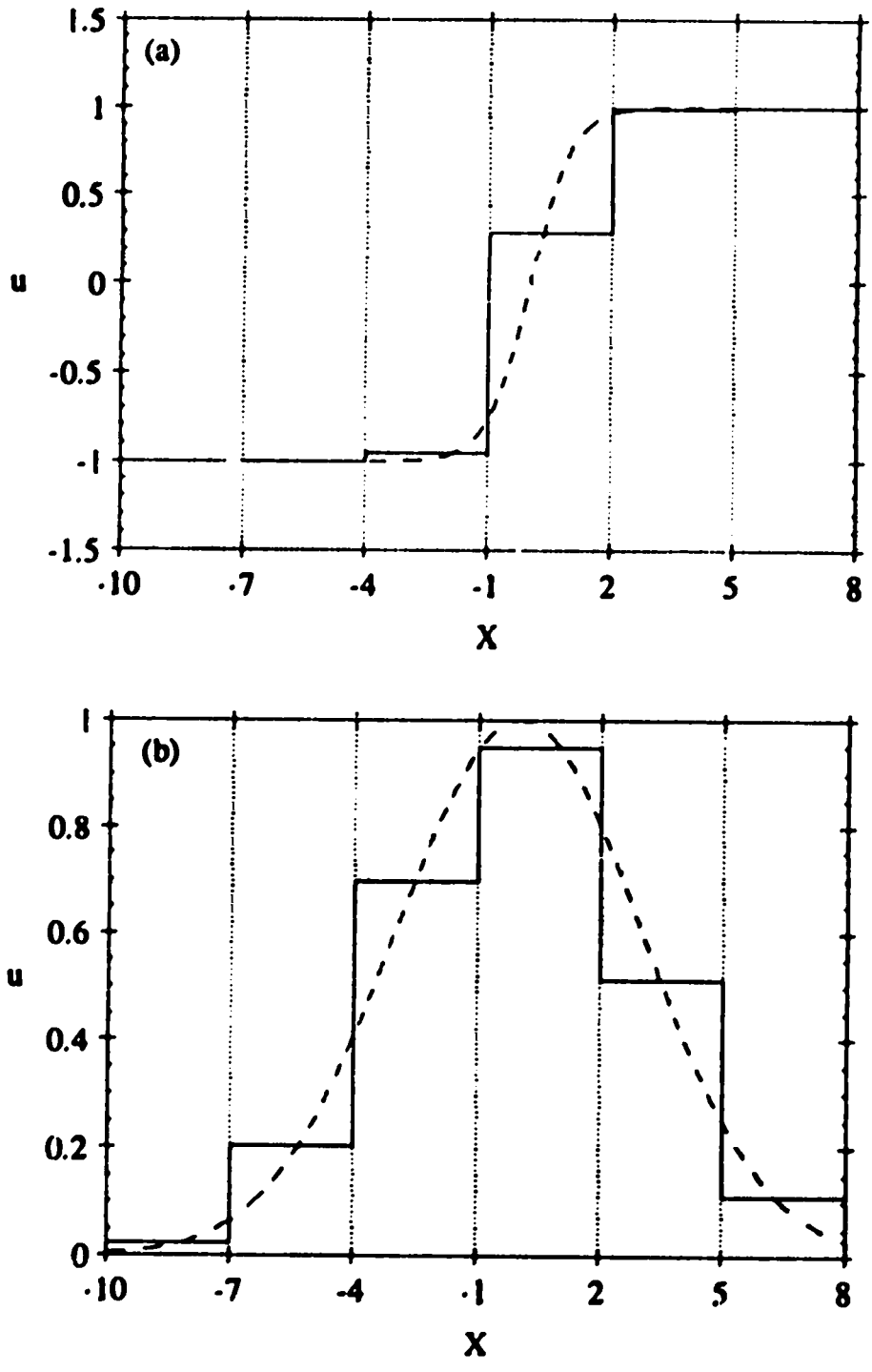


Figure 9.2: The reconstruction of the test functions by Godunov's method. The exact functions are given by the dashed lines. The grid on the plot denotes the computational grid.

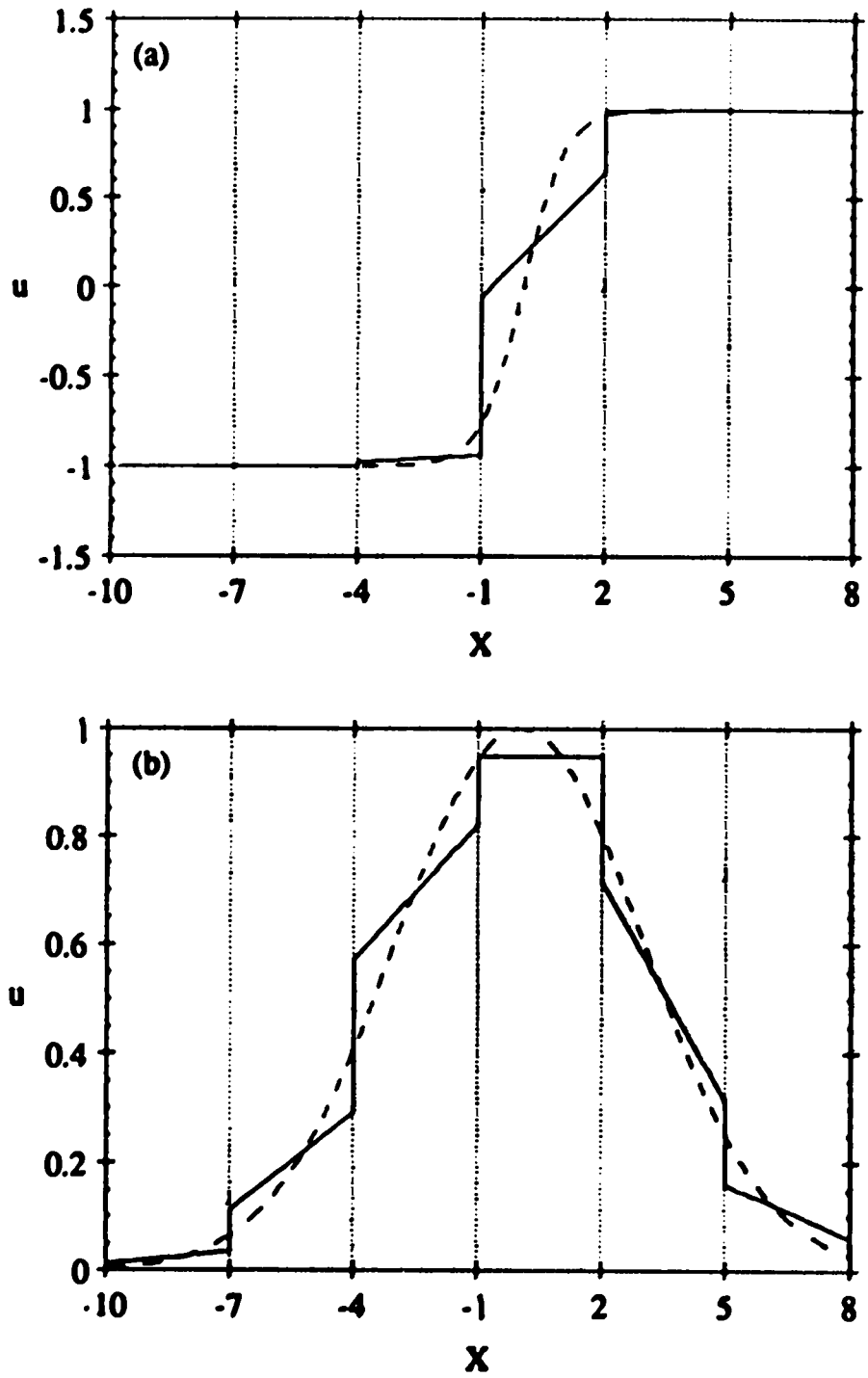


Figure 9.3: The reconstruction of the test functions by a second-order HOG method with the minmod limiter.

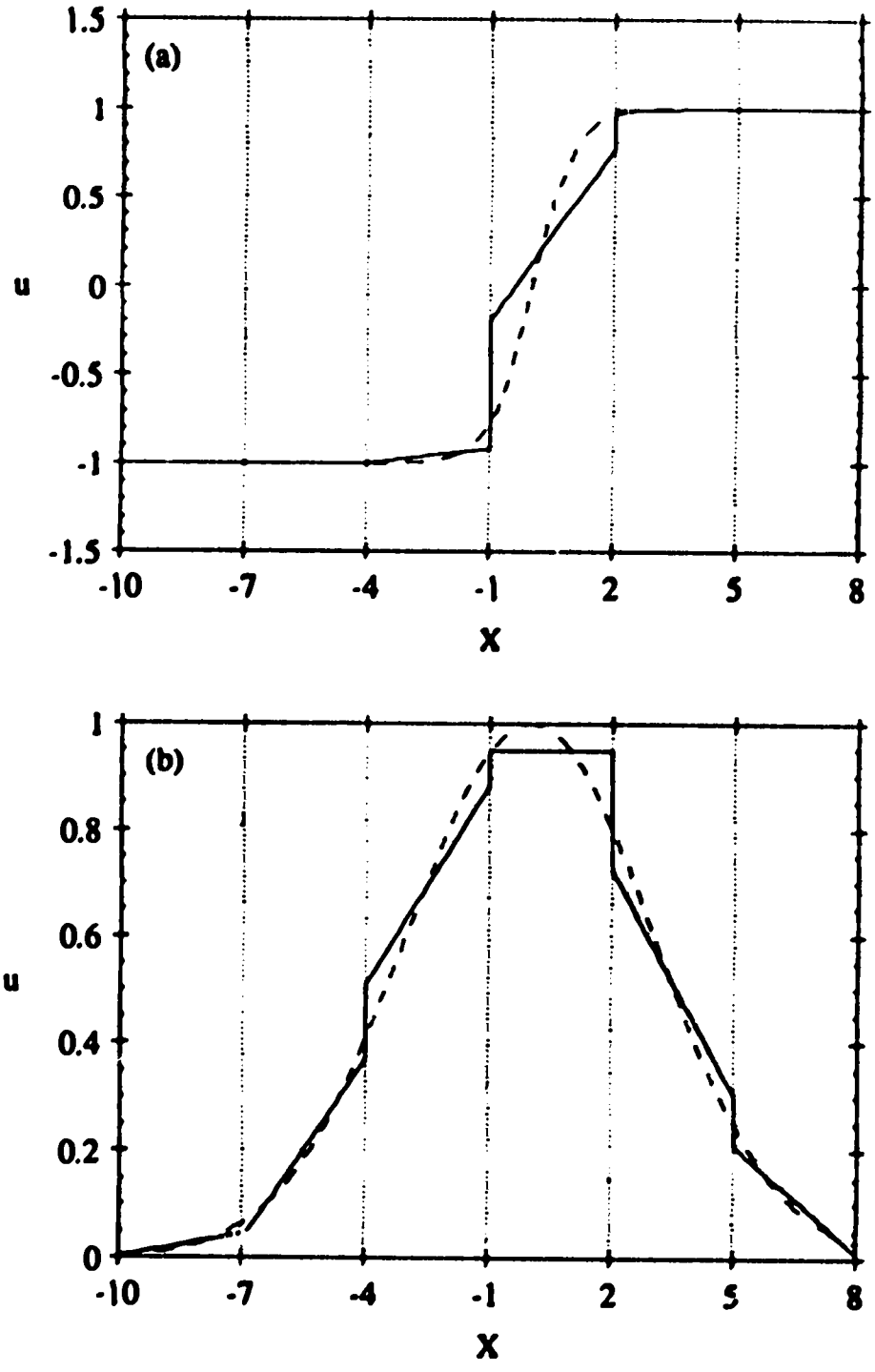


Figure 9.4: The reconstruction of the test functions by a second-order HOG method with the centered limiter.

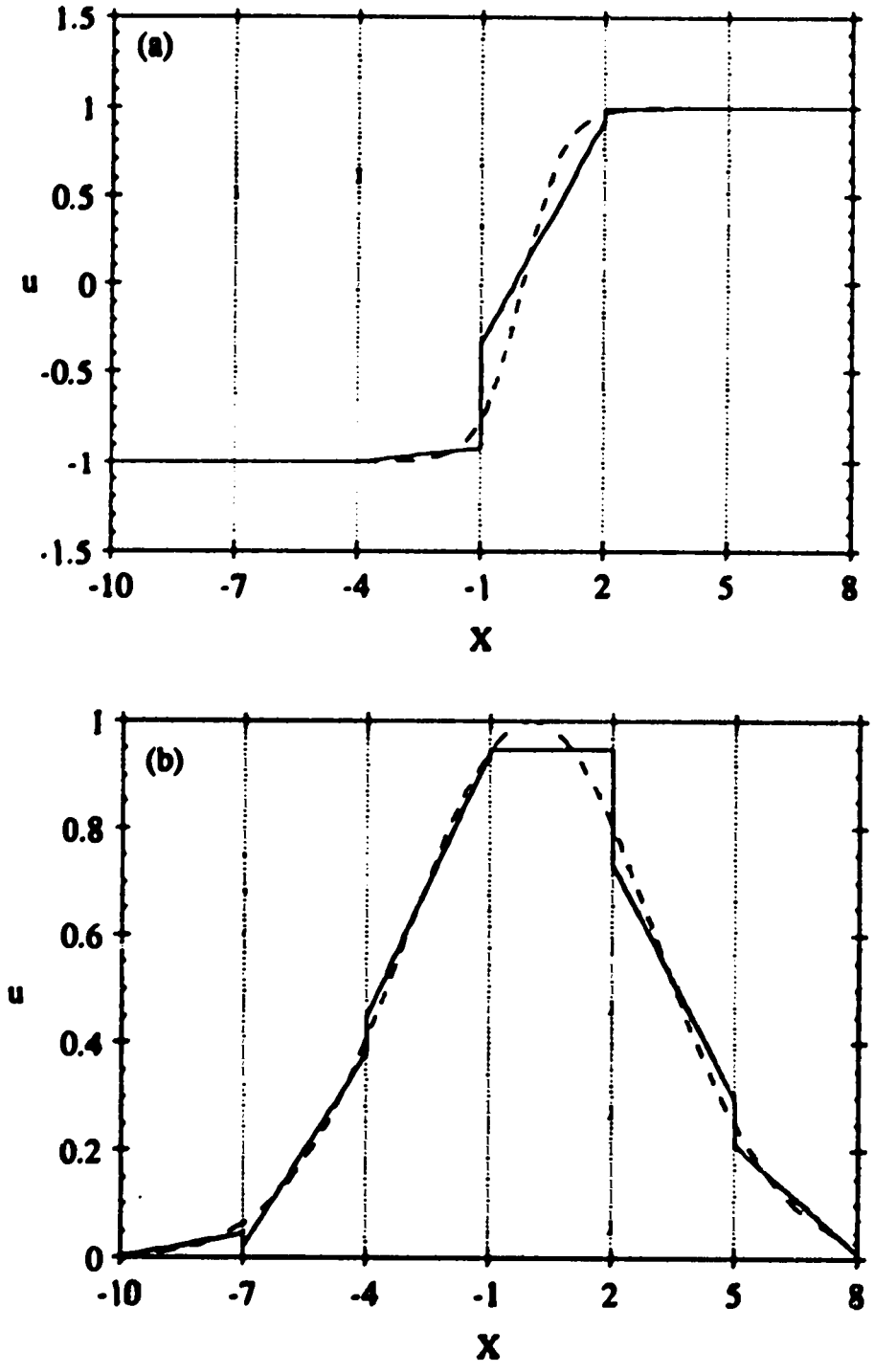


Figure 9.5: The reconstruction of the test functions by a second-order HOG method with the superbee limiter.

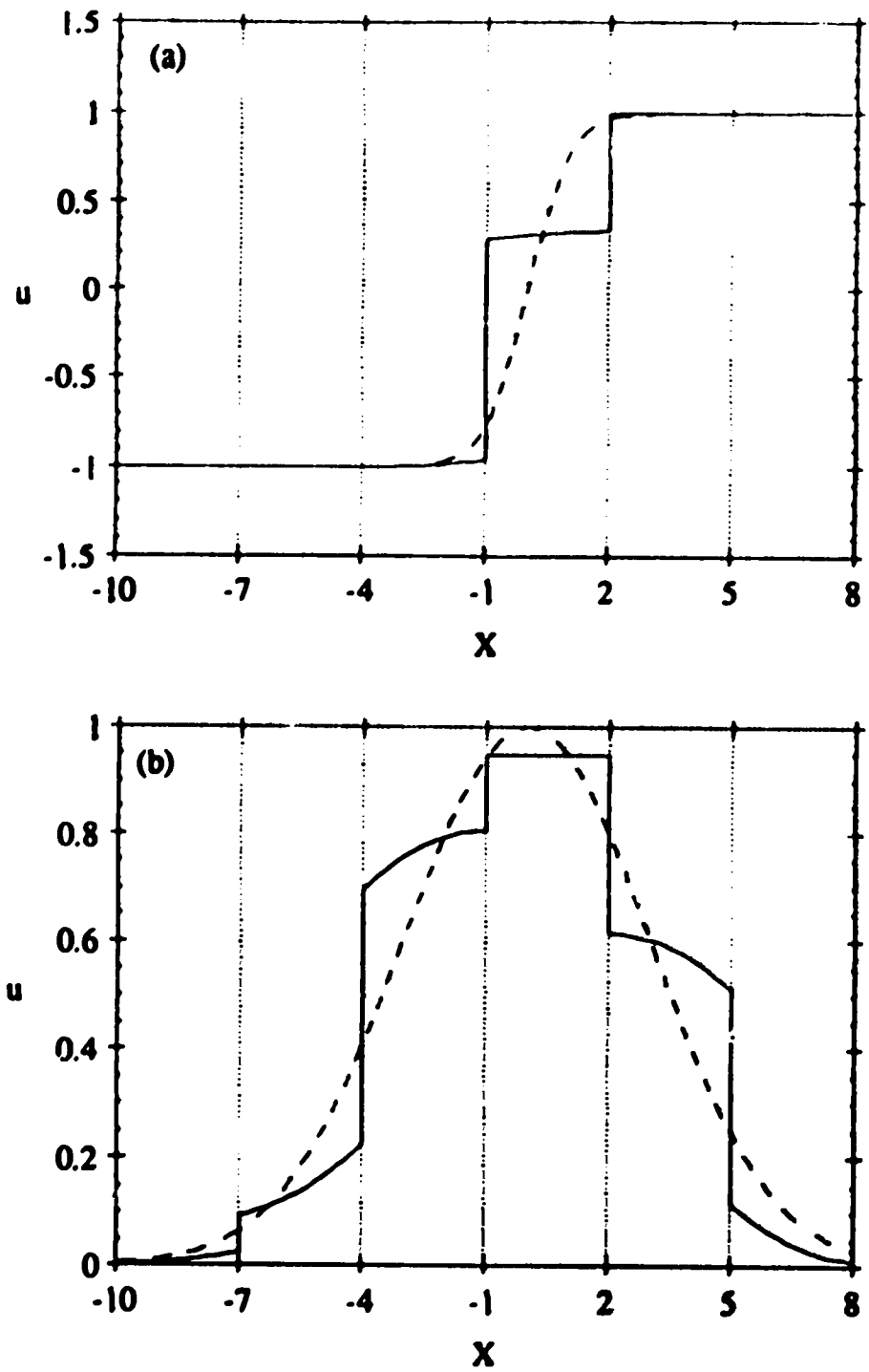


Figure 9.6: The reconstruction of the test functions by a MUSCL method with the three argument centered limiter.

Table 9.2: The type of scheme produced for various values of κ with the quadratic HOG reconstruction.

κ	Scheme
-1	one-sided, second-order
0	upwind, second-order
1/2	upwind, third-order
1	centered, second-order

Another interesting cell-average form can be found through imposing the constraint on the symmetric HOG scheme of giving a cell-average reconstruction. The scheme that results is

$$P_j(x) = u_j - \frac{\Delta x\left(\dot{s}_{j+\frac{1}{2}} - \dot{s}_{j-\frac{1}{2}}\right)}{8} + \begin{cases} \dot{s}_{j+\frac{1}{2}}(x - x_j) \; ; x \in \left[x_j, x_{j+\frac{1}{2}}\right] \\ \dot{s}_{j-\frac{1}{2}}(x - x_j) \; ; x \in \left[x_{j-\frac{1}{2}}, x_j\right] \end{cases} . \qquad (9.7)$$

Caution must be used with this scheme with regard to retaining TVD properties. In general upwind limiters do not produce a TVD scheme because the information from the upwind limiter is passed downwind via the correction term that assures the cell-average property, but a three argument centered limiter does not have these difficulties.

As I did in the cell-average section, the point-value interpolants are tested. In both cases shown below, three argument centered limiters are used. In Fig. 9.7 the symmetric HOG method is shown and in Fig. 9.8, the quadratic HOG (κ = 1/2) method is shown. The three argument limiters are too diffuse to capture the discontinuity in the $\tanh(x)$ function. The figures also show how the interpolants are C^1 continuous at the cell edges. Both are roughly equivalent to the MUSCL method in accuracy.

9.4 Results

This section presents results using methods described in the previous sections. Results cover the resolution, accuracy, economy, and general quality of the solutions. In order to do this, three types of problems are examined: the scalar wave equation, Burgers' equation, and the Euler equation. The test problems are all discussed in Appendix A.

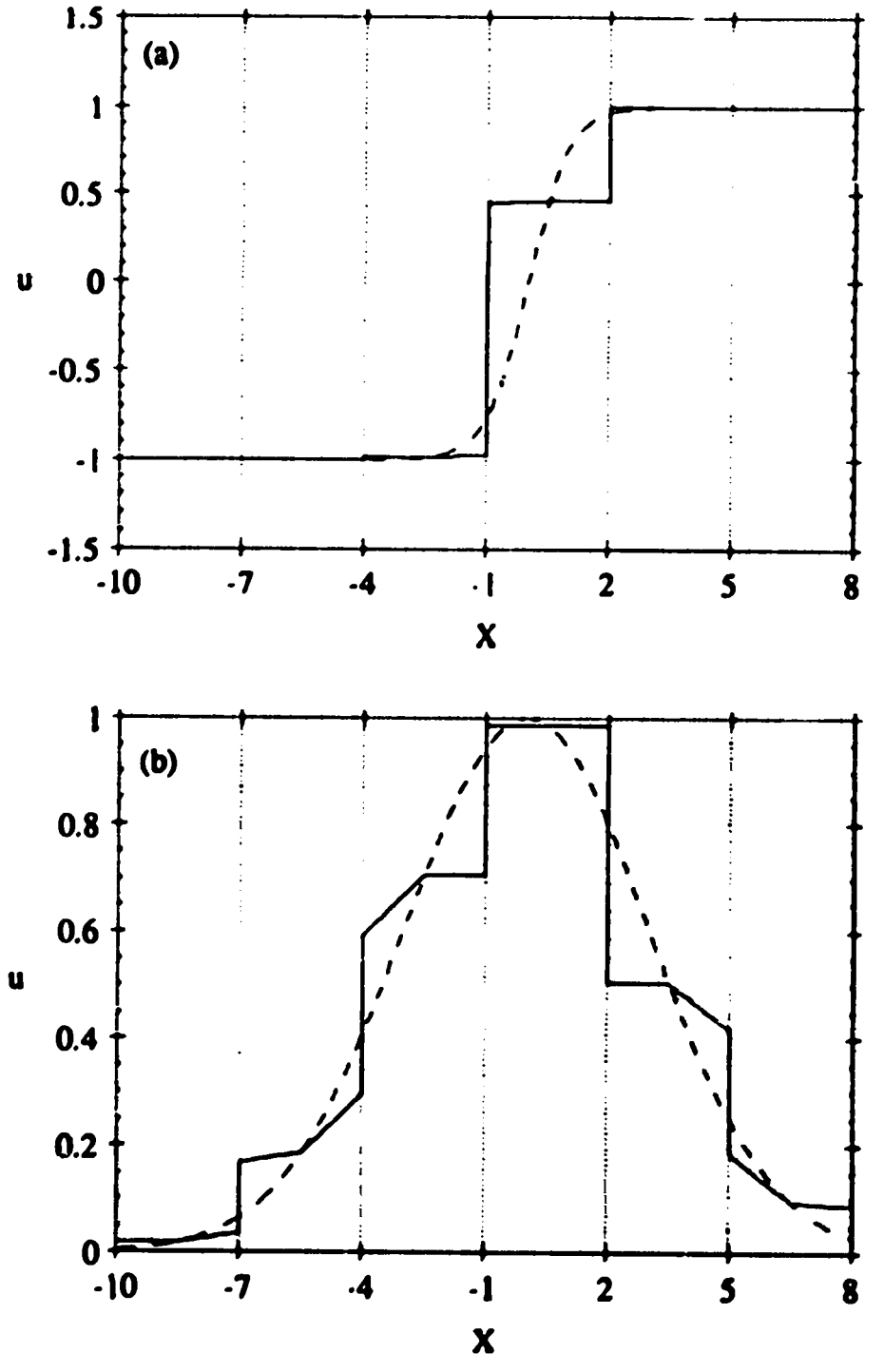


Figure 9.7: The reconstruction of the test functions by a symmetric HOG method with the three argument centered limiter.

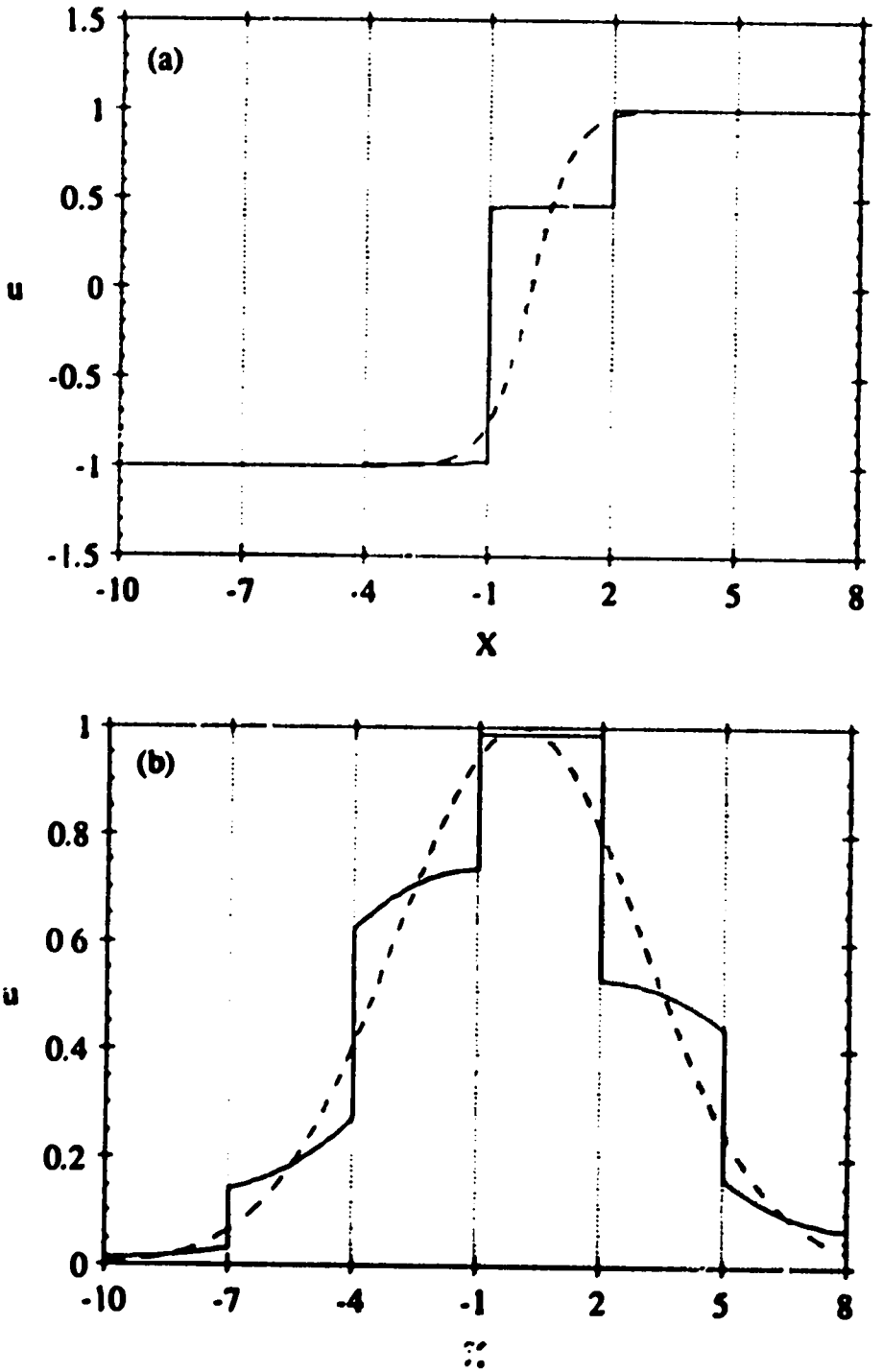


Figure 9.8: The reconstruction of the test functions by a quadratic HOG method with the three argument centered limiter.

Table 9.3: Sum of numerical viscous flux for the scalar wave equation test problems at $t = 250.0$.

Scheme	Sine Squared	Square
(3.12a) upwind biased	30.84	40.95
(9.7) upwind biased	31.36	43.27
(3.12a) symmetric	54.67	61.80
(9.7) symmetric	53.83	60.81
(9.6) $\kappa = 1/2$ minmod	53.98	60.96
(9.3) $\kappa = 1/3$ minmod	53.79	60.79
(9.6) $\kappa = 1/2$ centered	13.65	28.69
(9.3) $\kappa = 1/3$ centered	12.97	28.29
(9.6) $\kappa = 1/2$ MUSCL	30.25	39.51
(9.3) $\kappa = 1/3$ MUSCL	30.52	40.06

9.4.1 Scalar Wave Equation

In addition to a comparison of the qualitative appearance of the results, several quantitative measures of algorithmic performance are used: the peak values in the solutions, the total variation of the solution at the end of the test and a measure of numerical viscosity. The measure of numerical viscosity is made by a technique described in a general sense in [30]. This idea was expanded on by the author in Chapter 8. The gist of the technique is to compare the numerical fluxes of a high-order technique with that of the Lax-Wendroff method and denote the difference as numerical viscosity. The results for various methods using this approach are shown in Table 9.3. For the schemes that are TVD for both construction techniques, the cell-average reconstruction carries less numerical viscosity, but when the schemes are not TVD, cell-average reconstruction is more viscous. This general conclusion is born out by other depictions of the data.

The results shown in Table 9.4 show that, in general, the two methods of reconstruction yield similar results for similar schemes. Except for the upwind-biased Lax-Wendroff type scheme, these results are consistent with the measure of numerical viscosity. Figure 9.9 shows the excellent results obtained with the upwind-biased Lax-Wendroff TVD scheme. Making this scheme a cell-average reconstruction destroys its TVD property and makes the results (shown in Fig. 9.10) quite poor although the maximum values are improved.

Table 9.4: Maximum profile values for the scalar wave equation test problems at $t = 250.0$.

Scheme	Sine Squared	Square
(3.12a) upwind biased	0.9197	0.7108
(9.7) upwind biased	0.9481	0.7598
(3.12a) symmetric	0.8717	0.6037
(9.7) symmetric	0.8689	0.6030
(9.6) $\kappa = 1/2$ minmod	0.8690	0.6032
(9.3) $\kappa = 1/3$ minmod	0.8689	0.6031
(9.6) $\kappa = 1/2$ centered	0.9602	0.7795
(9.3) $\kappa = 1/3$ centered	0.9603	0.7799
(9.6) $\kappa = 1/2$ MUSCL	0.9394	0.7519
(9.3) $\kappa = 1/3$ MUSCL	0.9334	0.7487

In the case of the symmetric HOC scheme, the method remains TVD after its transformation to a cell-average reconstruction. Figures 9.11 and 9.12 show the results obtained with these methods. The point-value reconstruction gives slightly higher resolution and less viscosity, but the cell-average reconstruction results in a solution with better symmetry properties.

As shown in Figs. 9.13-9.16 these results carry over to the quadratic reconstructions using the minmod limiter, but not to the centered limiter, which slightly favors the cell-average reconstruction from every perspective. This included the qualitative appearance of the solutions. The classic-MUSCL (nonTVD) solutions are similar, but the results do not favor the cell-average reconstruction for the square wave. In this case the oscillations are worse.

9.4.2 Burgers' Equation

This section of the chapter discusses the order of accuracy of the reconstructions and their subsequent solutions.

Table 9.5 shows the rates of convergence obtained with some of these methods when the solution is smooth. In every case, the rates of convergence obtained with the point-value reconstruction are superior, in some cases by quite a margin. This is equally true for the solutions after a shock has formed. Table 9.6 shows this quite clearly and in some cases the disparity in performance is quite profound.

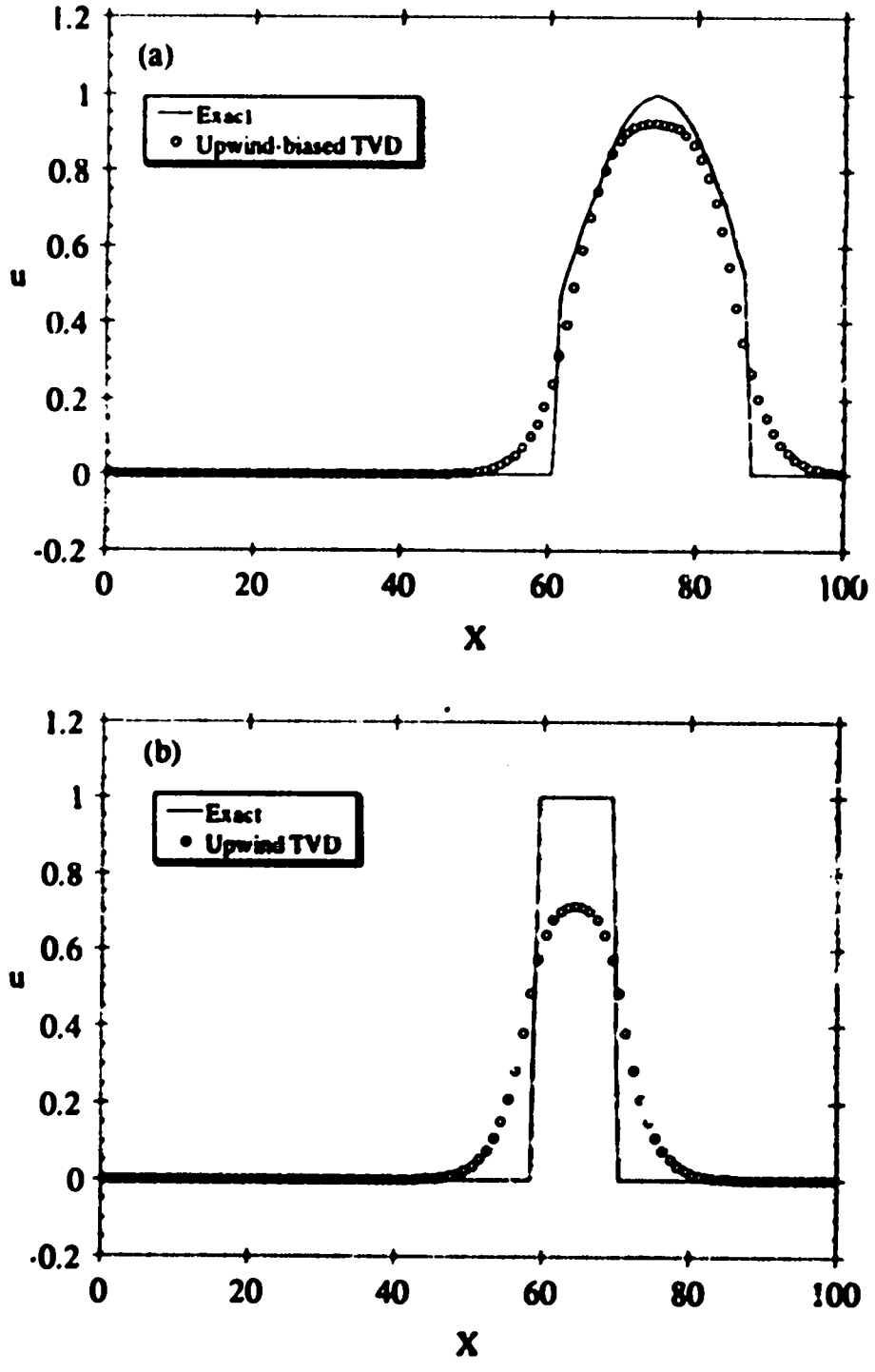


Figure 9.9: The solution to the scalar wave equation by an upwind-biased Lax-Wendroff TVD method.

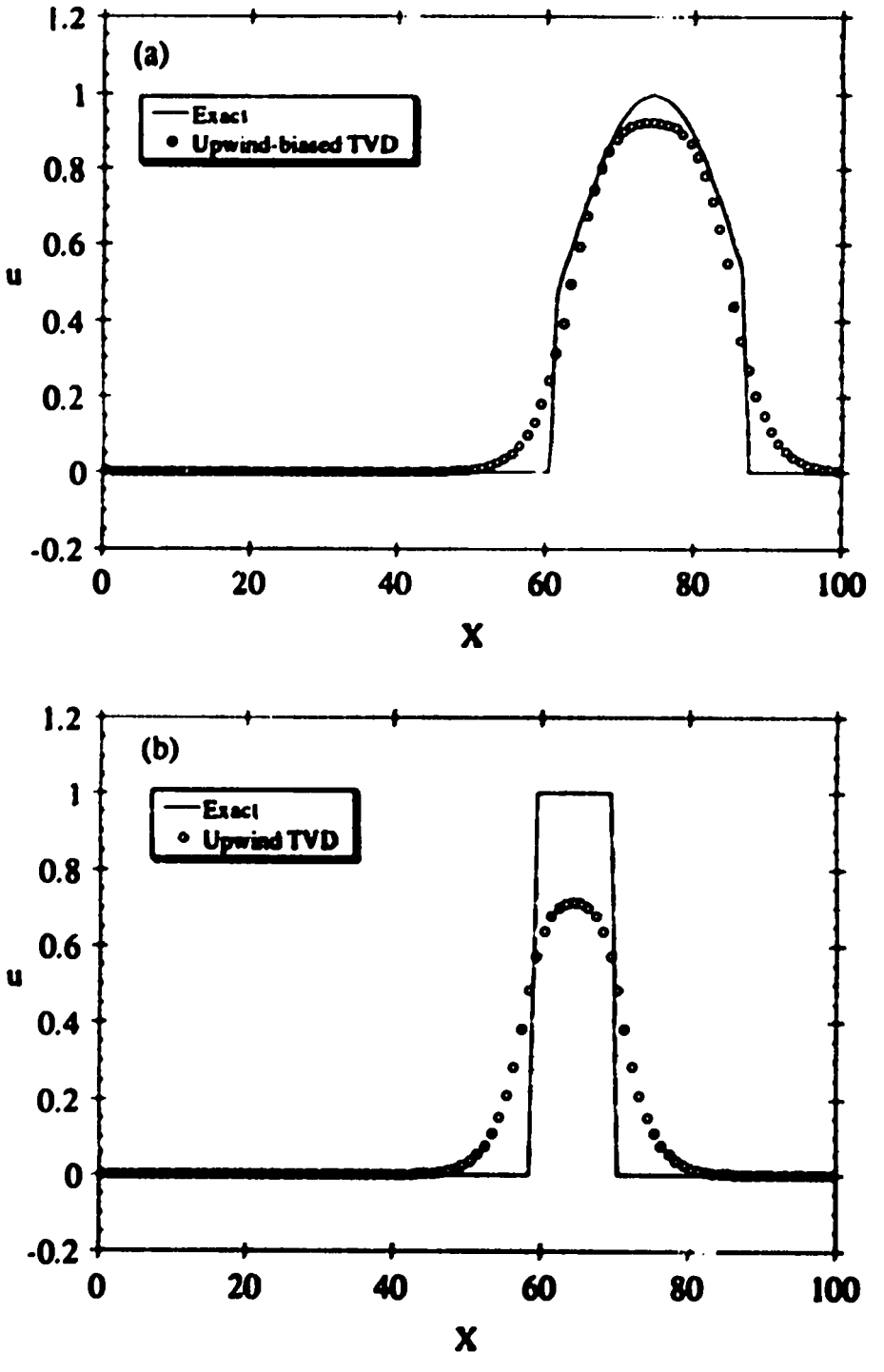


Figure 9.10: The solution to the scalar wave equation by an upwind-biased Lax-Wendroff TVD method with a cell-average correction.

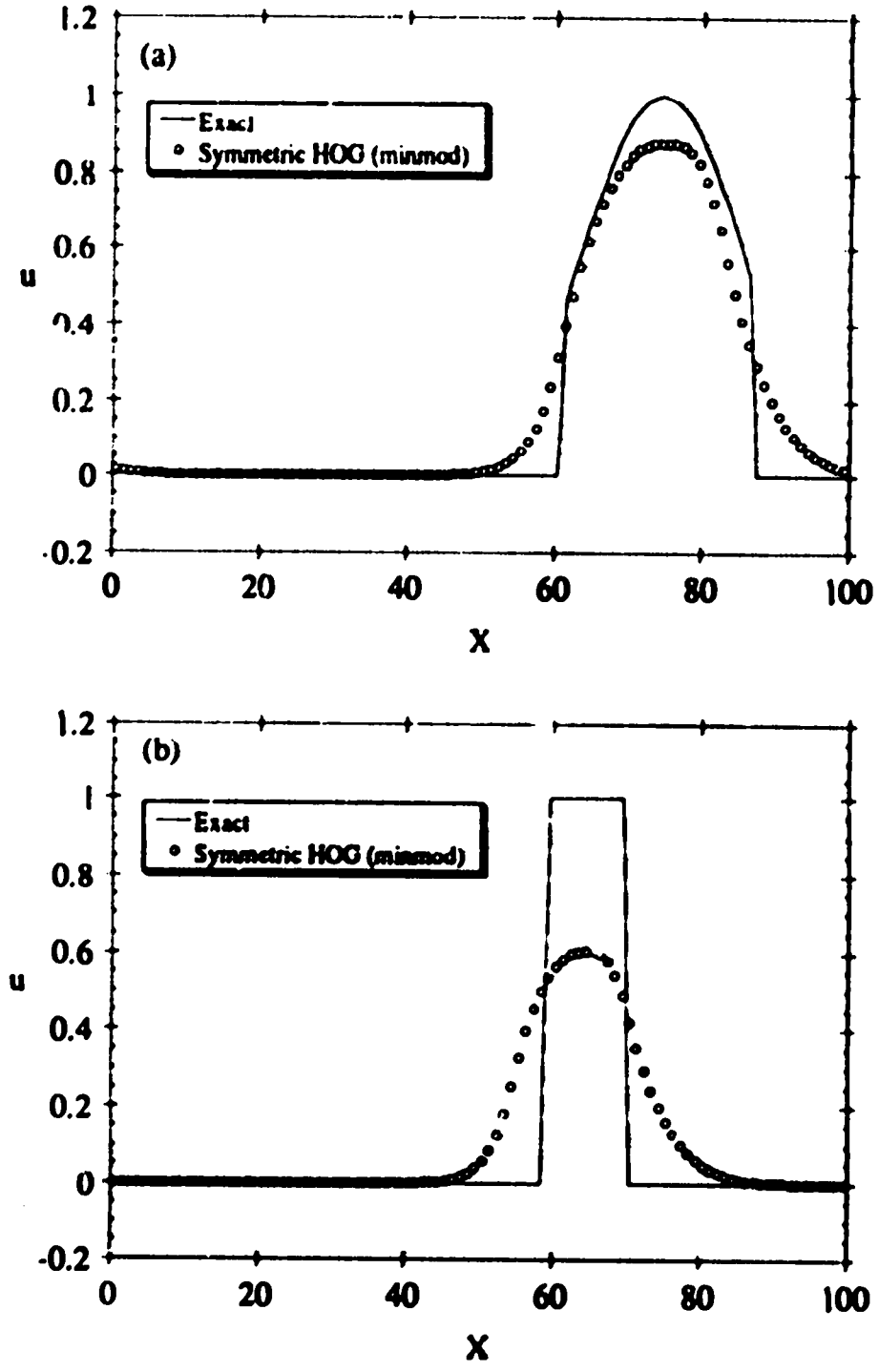


Figure 9.11: The solution to the scalar wave equation by a symmetric HOG method.

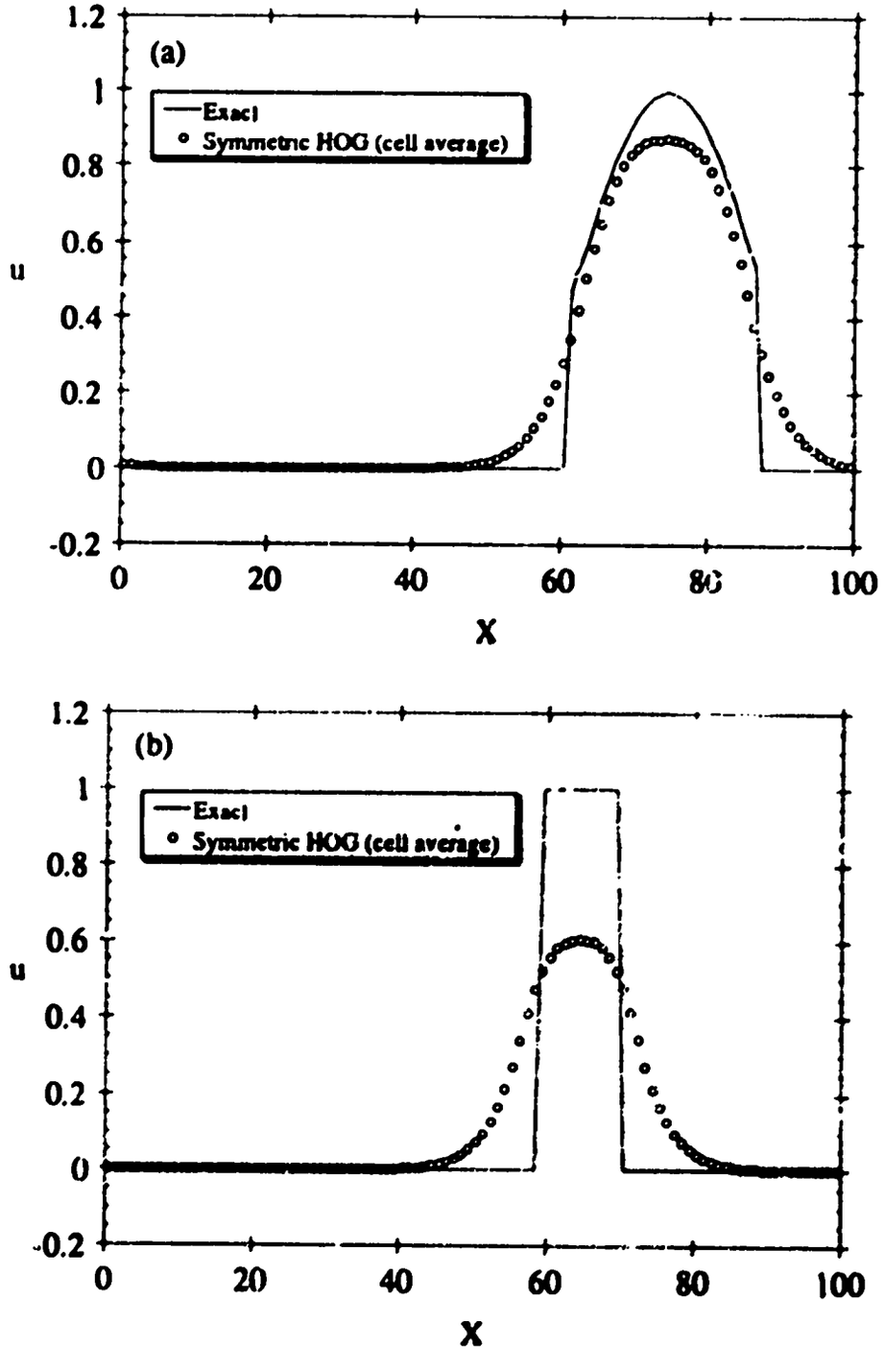


Figure 9.12: The solution to the scalar wave equation by a symmetric HOG method with a cell-average correction.

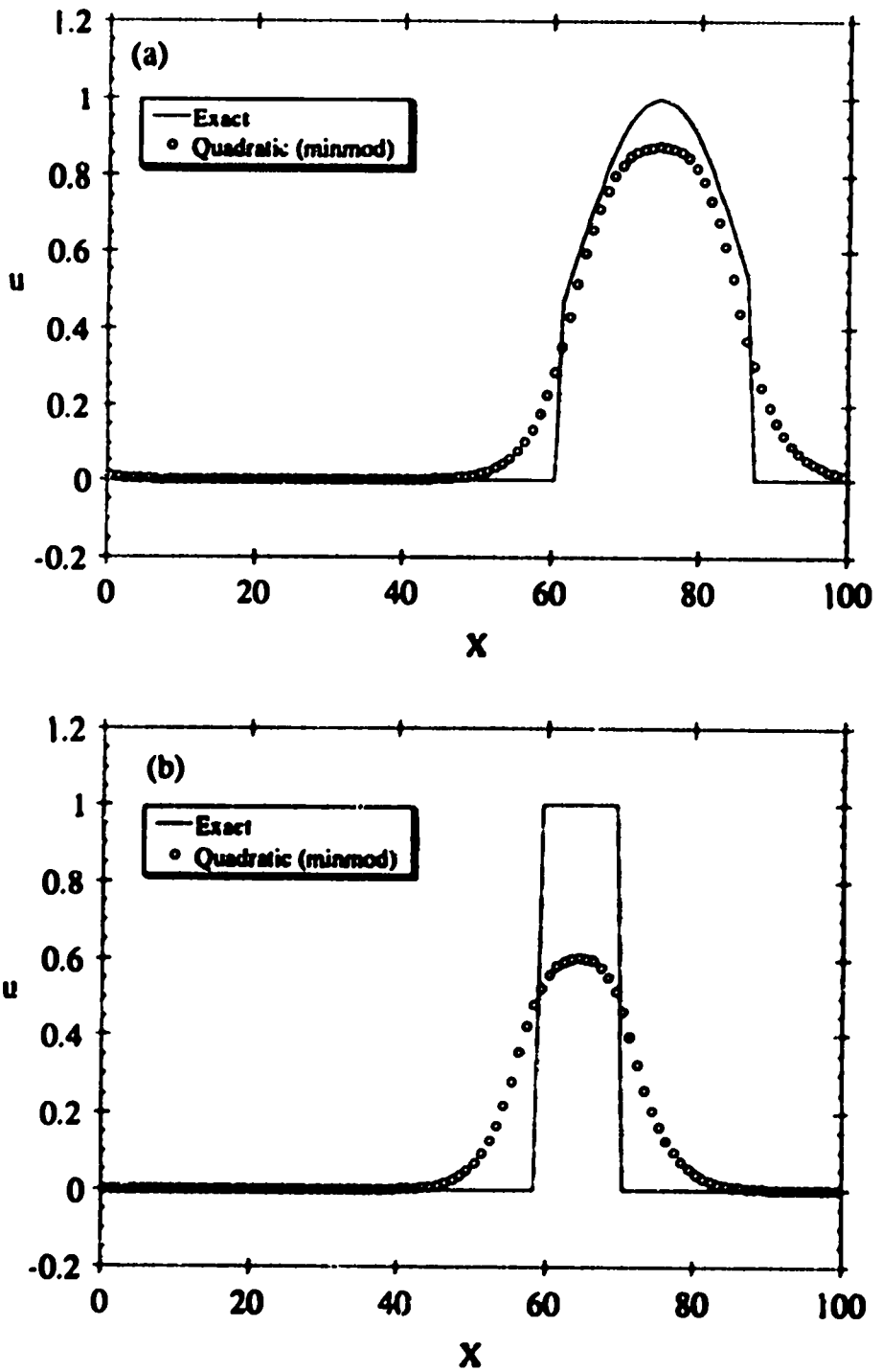


Figure 9.13: The solution to the scalar wave equation by a quadratic Taylor polynomial based HOG method with a minmod limiter.

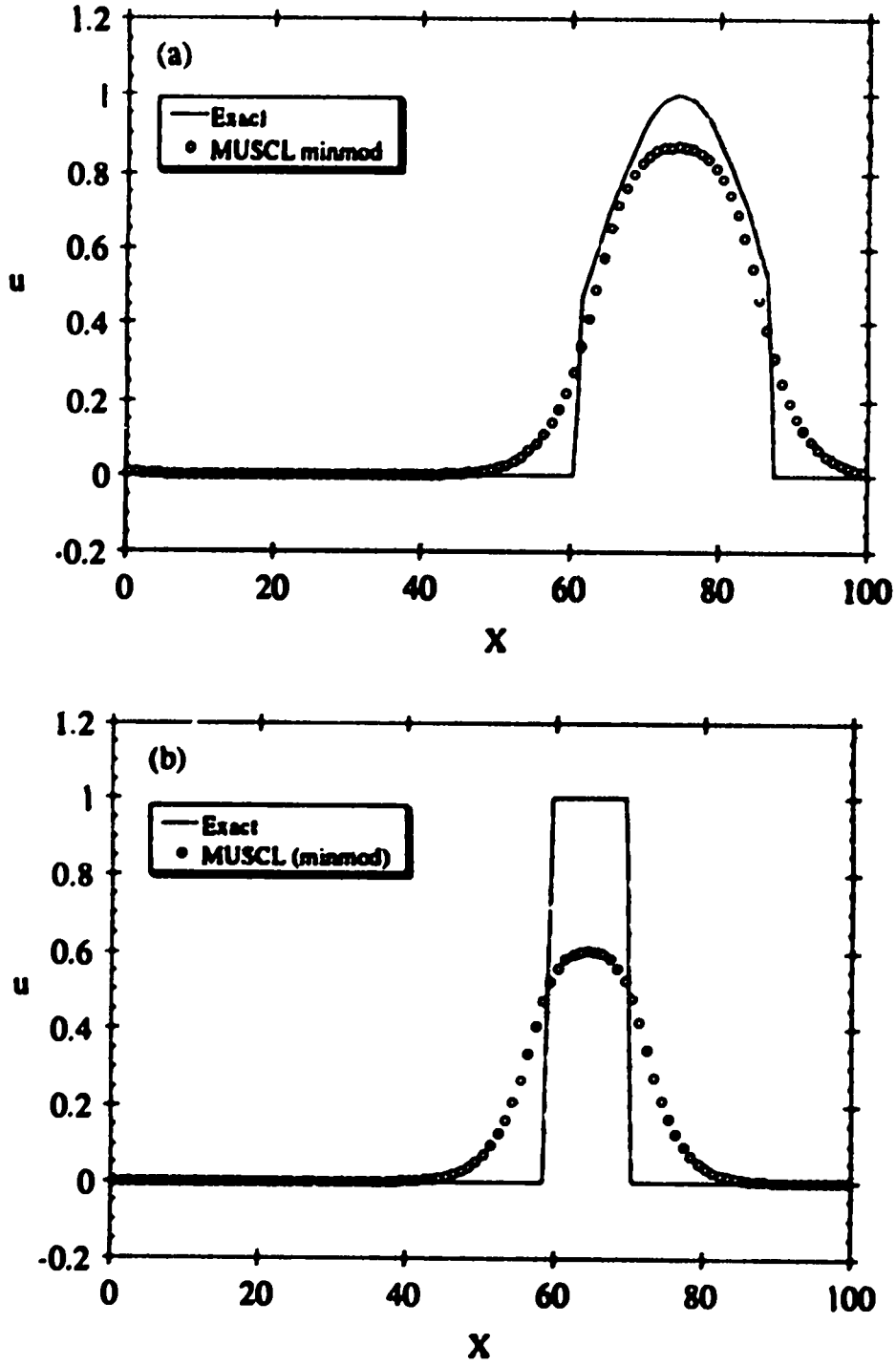


Figure 9.14: The solution to the scalar wave equation by a quadratic Legendre polynomial based HOG method with a minmod limiter.

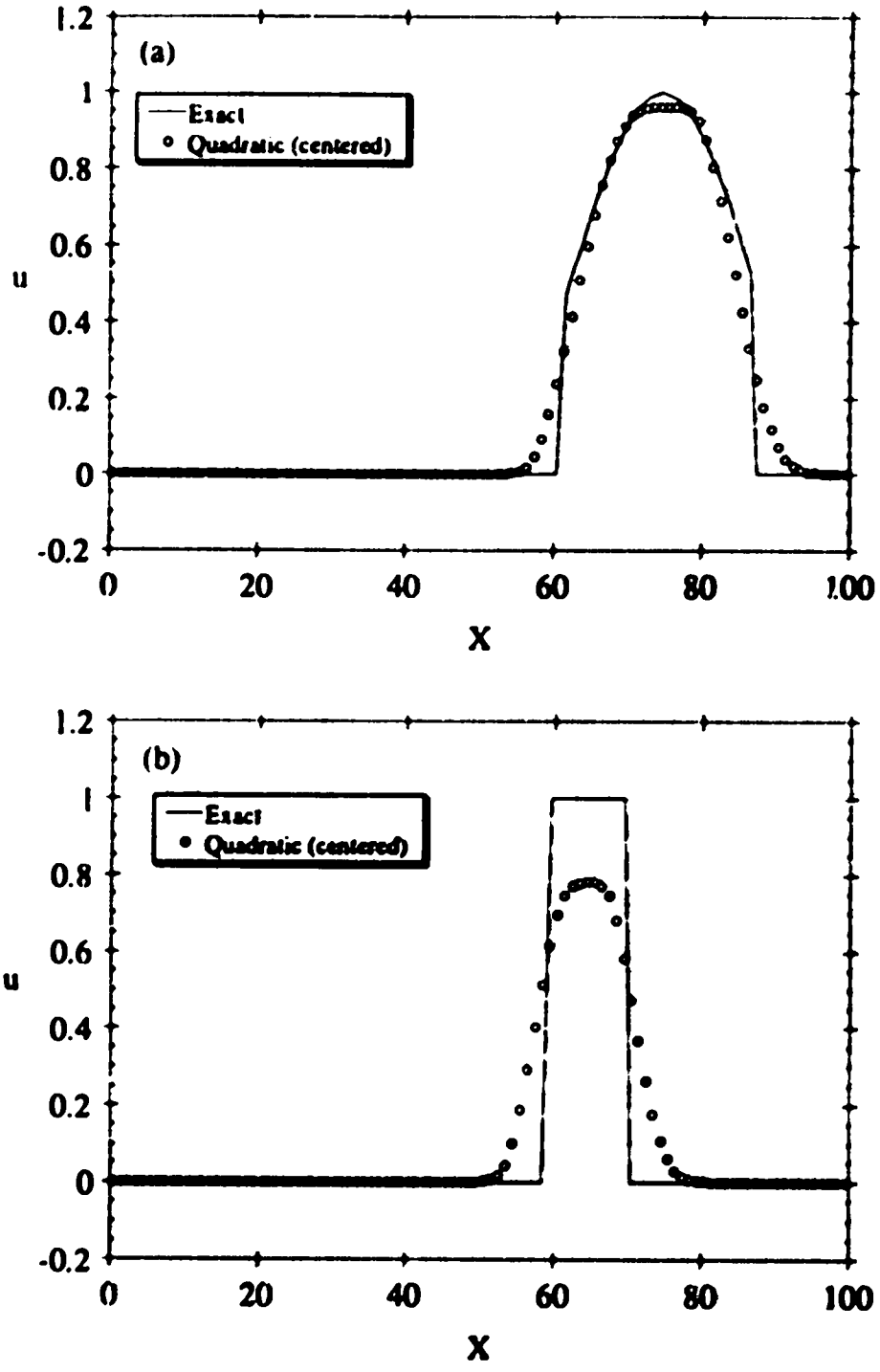


Figure 9.15: The solution to the scalar wave equation by a quadratic Taylor polynomial based HOG method with a centered limiter.

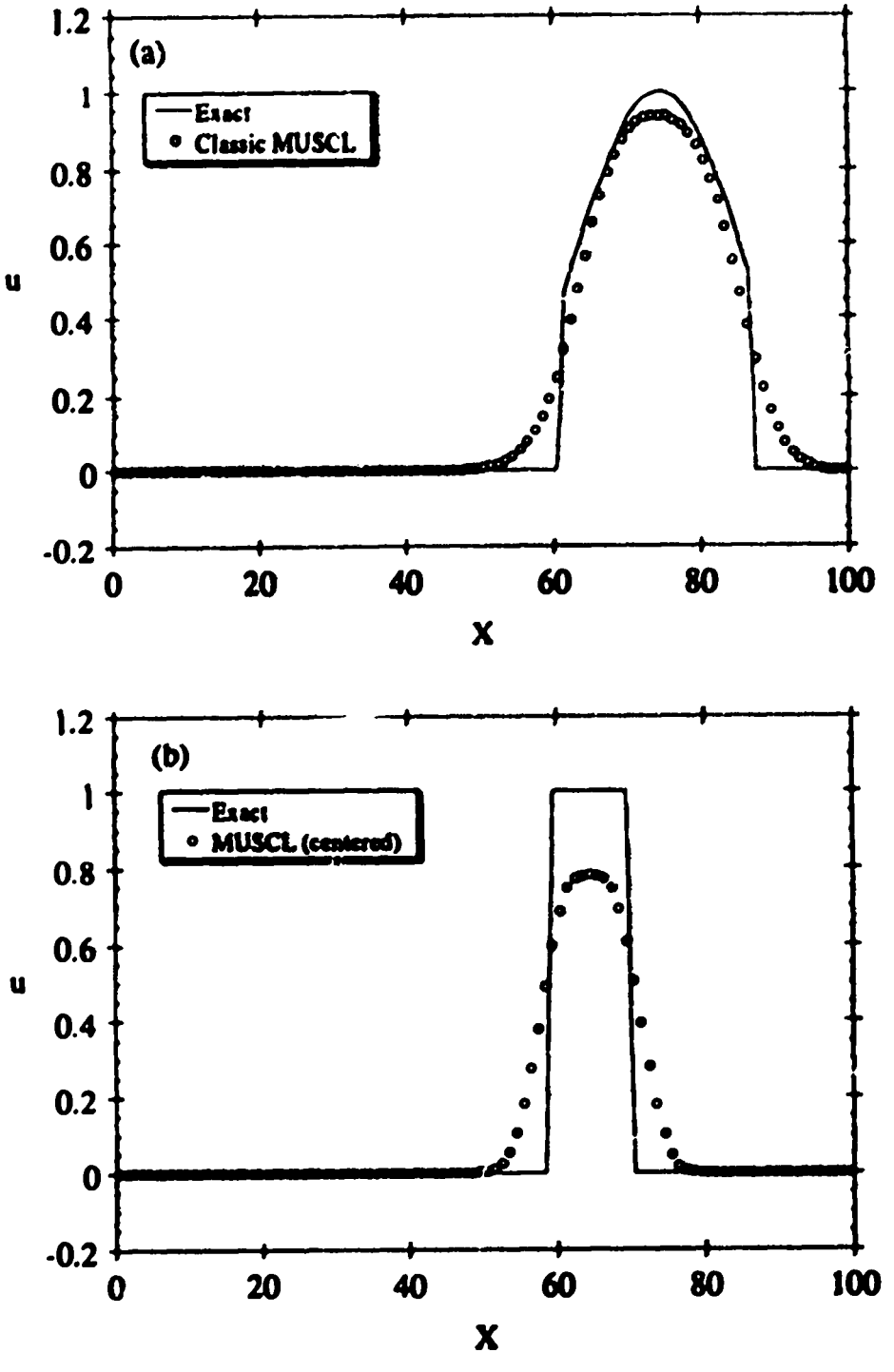


Figure 9.16: The solution to the scalar wave equation by a quadratic Legendre polynomial based HOG method with a centered limiter.

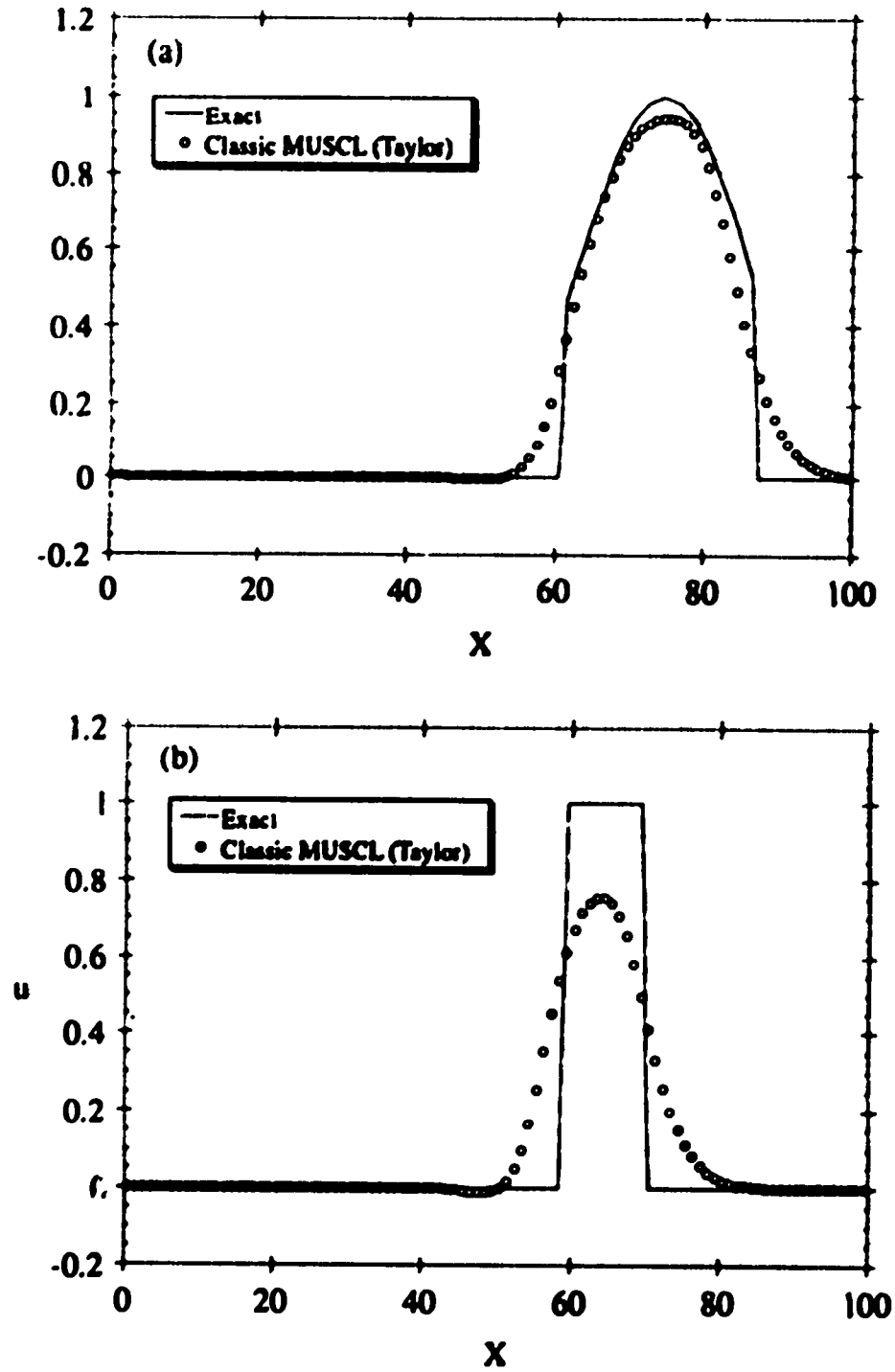


Figure 9.17: The solution to the scalar wave equation by a Taylor polynomial based classic MUSCL scheme.

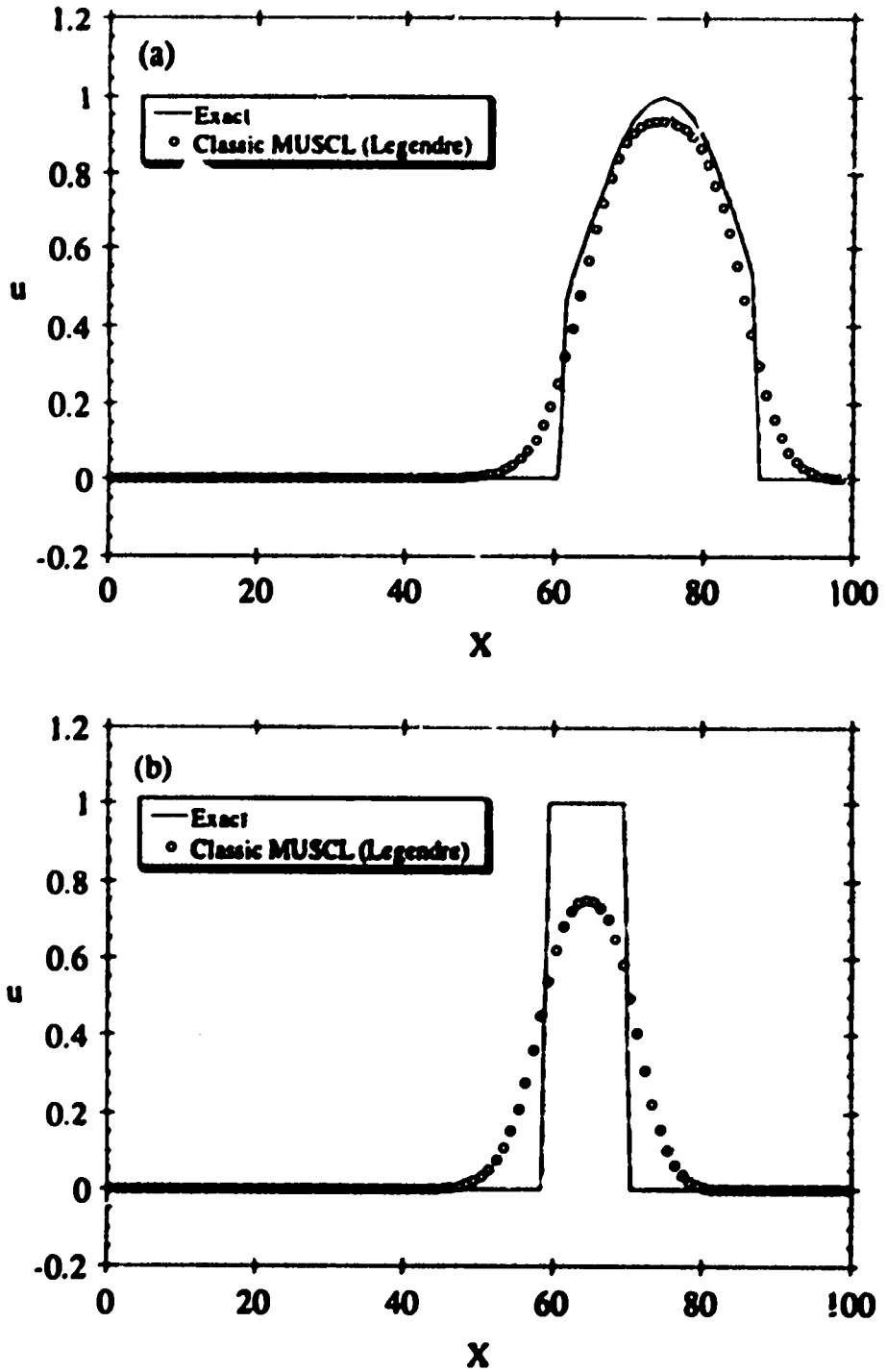


Figure 9.18: The solution to the scalar wave equation by a Legendre polynomial based classic MUSCL scheme.

Table 9.5: The order of convergence in several norms for various schemes for Burgers' equation at $t = 0.2$ when the solution is smooth.

Scheme	L_1	L_2	L_∞
(9.2)	2.16	2.17	1.95
(3.12a) upwind biased	2.16	2.17	1.95
(3.12a) symmetric	2.13	1.86	1.32
(3.12a) symmetric	1.87	1.60	1.23
(9.6) $\kappa = 1/2$ TVD	2.11	1.83	1.28
(9.3) $\kappa = 1/3$ TVD	2.02	1.74	1.22
(9.6) $\kappa = 1/2$ MUSCL	2.05	1.72	1.15
(9.3) $\kappa = 1/3$ MUSCL	1.88	1.57	1.11

Table 9.6: The order of convergence in several norms for various schemes for Burgers' equation at $t = 1.0$ when the solution has a shock.

Scheme	L_1	L_2	L_∞
(9.2)	1.52	1.10	0.61
(3.12a) upwind biased	1.52	1.10	0.61
(3.12a) symmetric	1.53	1.00	0.47
(9.7) symmetric	0.71	0.58	0.36
(9.6) TVD $\kappa = 1/2$	1.61	1.05	0.53
(9.3) TVD $\kappa = 1/3$	1.60	1.07	0.56
(9.6) MUSCL $\kappa = 1/2$	1.43	1.02	0.54
(9.3) MUSCL $\kappa = 1/3$	0.98	0.78	0.53

Table 9.7: L_1 norms for density and velocity in Sod's problem, including times for reconstruction for each solution.

Scheme	Density	Velocity	Times
(9.2)	5.81×10^{-3}	1.13×10^{-2}	0.97
(3.12a) upwind biased	5.86×10^{-3}	1.15×10^{-2}	0.94
(9.7) upwind biased	8.15×10^{-3}	1.18×10^{-2}	0.93
(3.12a) symmetric	6.50×10^{-3}	1.02×10^{-2}	1.08
(9.7) symmetric	6.40×10^{-3}	9.99×10^{-3}	1.14
(9.6) TVD $\kappa = 1/2$	6.44×10^{-3}	1.01×10^{-2}	1.18
(9.3) TVD $\kappa = 1/3$	6.44×10^{-3}	1.01×10^{-2}	1.34

9.4.3 The Euler Equations

This section shows the performance of some of the methods discussed in this chapter on a system of conservation laws. As is common practice, the Euler equations are solved because of their great practical interest. It should demonstrate a "true" picture of each methods capabilities. For each of the methods used below, the density and velocity profiles are shown and the L_1 norms of these solutions are given.

The solutions are shown at $t = 20$. The solutions shown below use Roe's approximate Riemann solver and a characteristic variable based reconstruction [63, 200]. The TVD schemes using the three argument limiters employ the centered limiter for the nonlinear waves in equations and a superbee limiter for the linearly degenerate wave. For those methods using two argument limiters, the nonlinear waves use a van Leer limiter.

As shown in Figs. 9.19–9.25, the results obtained with these methods for systems of equations are all quite good. Each solution with the exception of the upwind-biased Lax-Wendroff type has a bump in the velocity solution at the end of the rarefaction wave. The solution obtained for the shock wave with this method is slightly better (two cells wide rather than three). Table 9.7 shows the methods' L_1 norms for density and velocity. In general, the results are similar here as well. For the upwind-biased Lax-Wendroff TVD methods, the cell-average form is noticeably inferior whereas the cell-average symmetric HOG method is superior to the corresponding point-value reconstruction. In general, the differences economy of use are inconsequential except for the classic MUSCL-Legendre formulation.

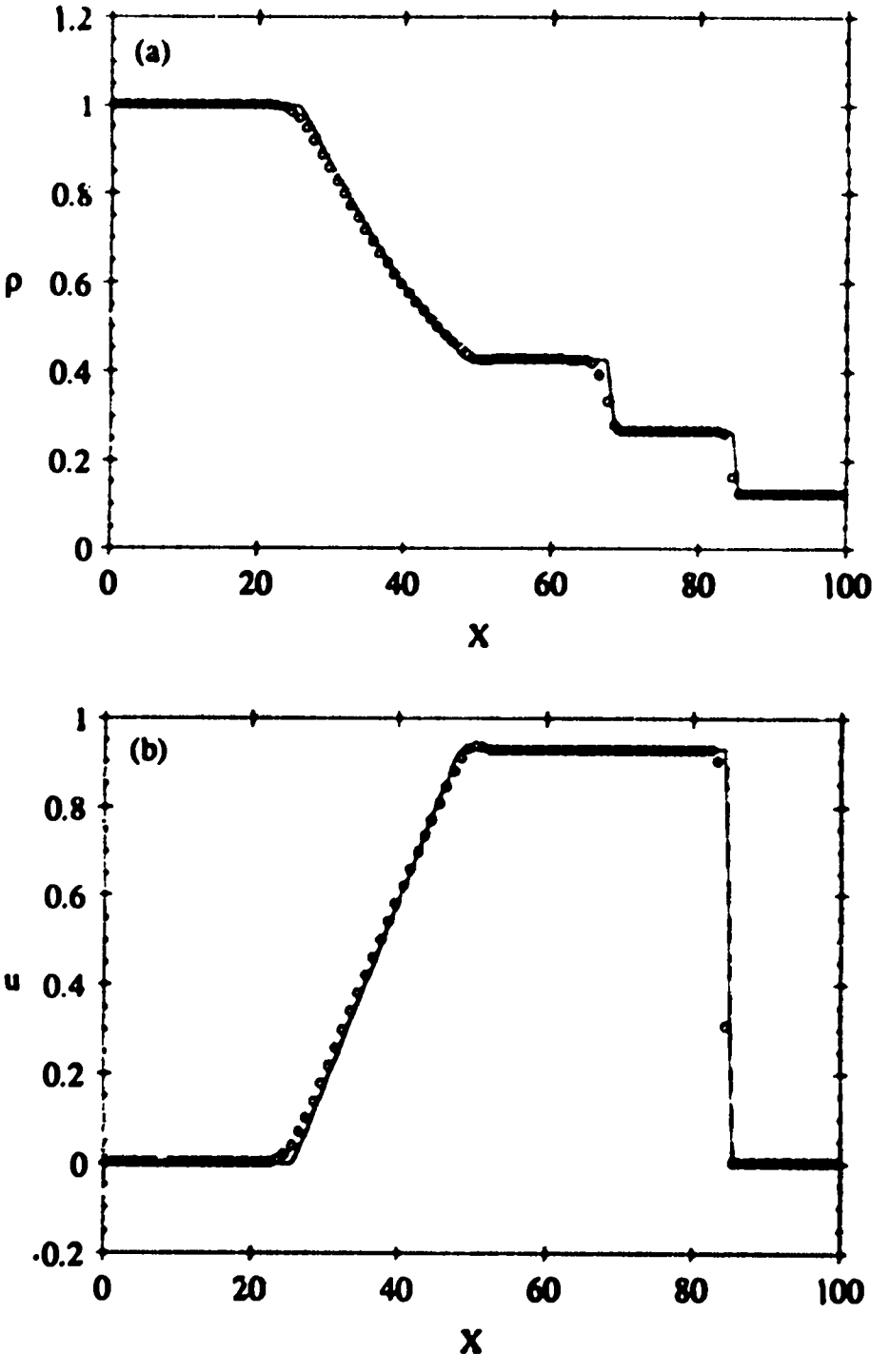


Figure 9.19: The density and velocity solutions to Sod's problem with a cell-average second-order HOG method.

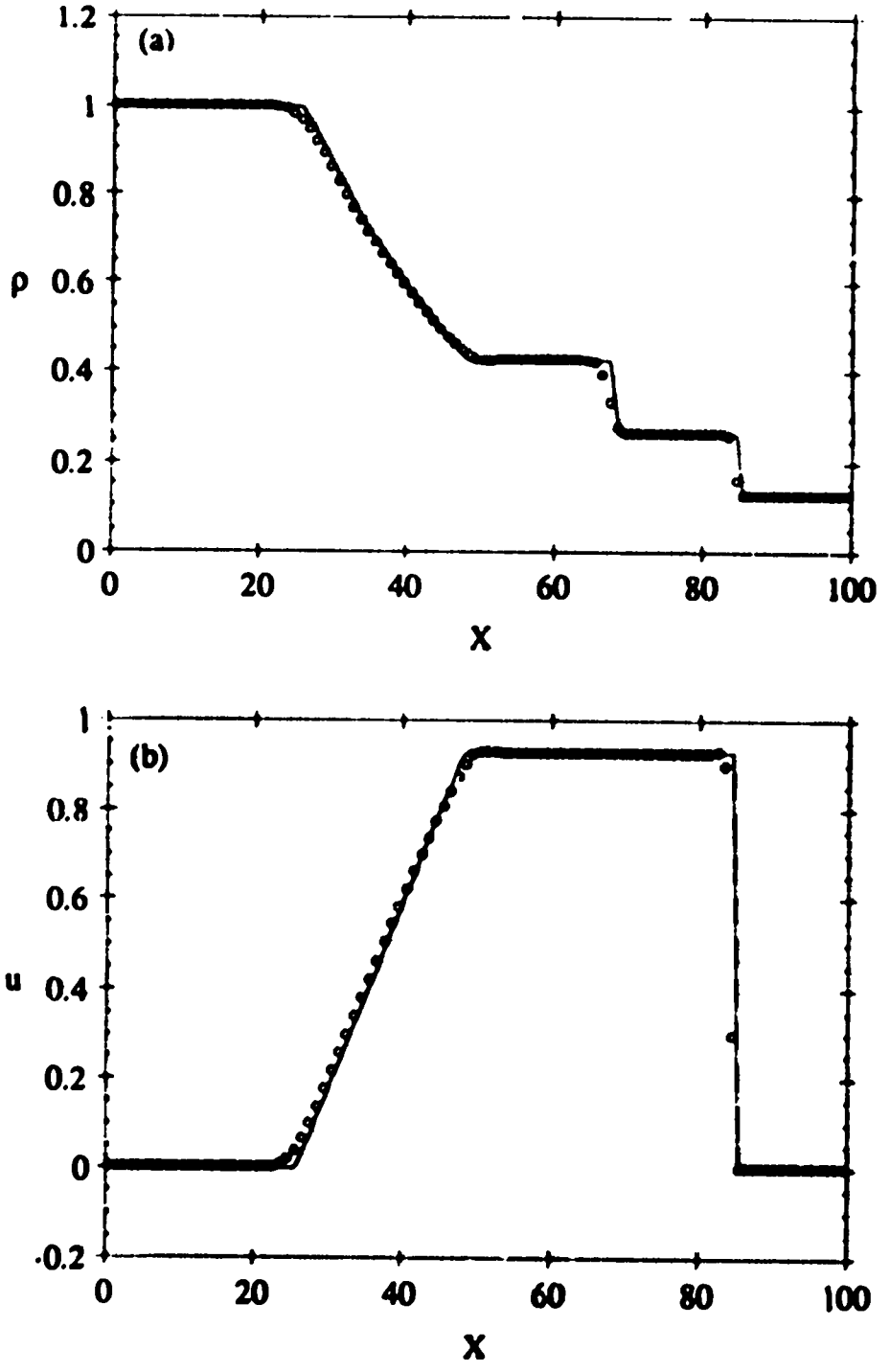


Figure 9.20: The density and velocity solutions to Sod's problem with an upwind-biased Lax-Wendroff TVD method

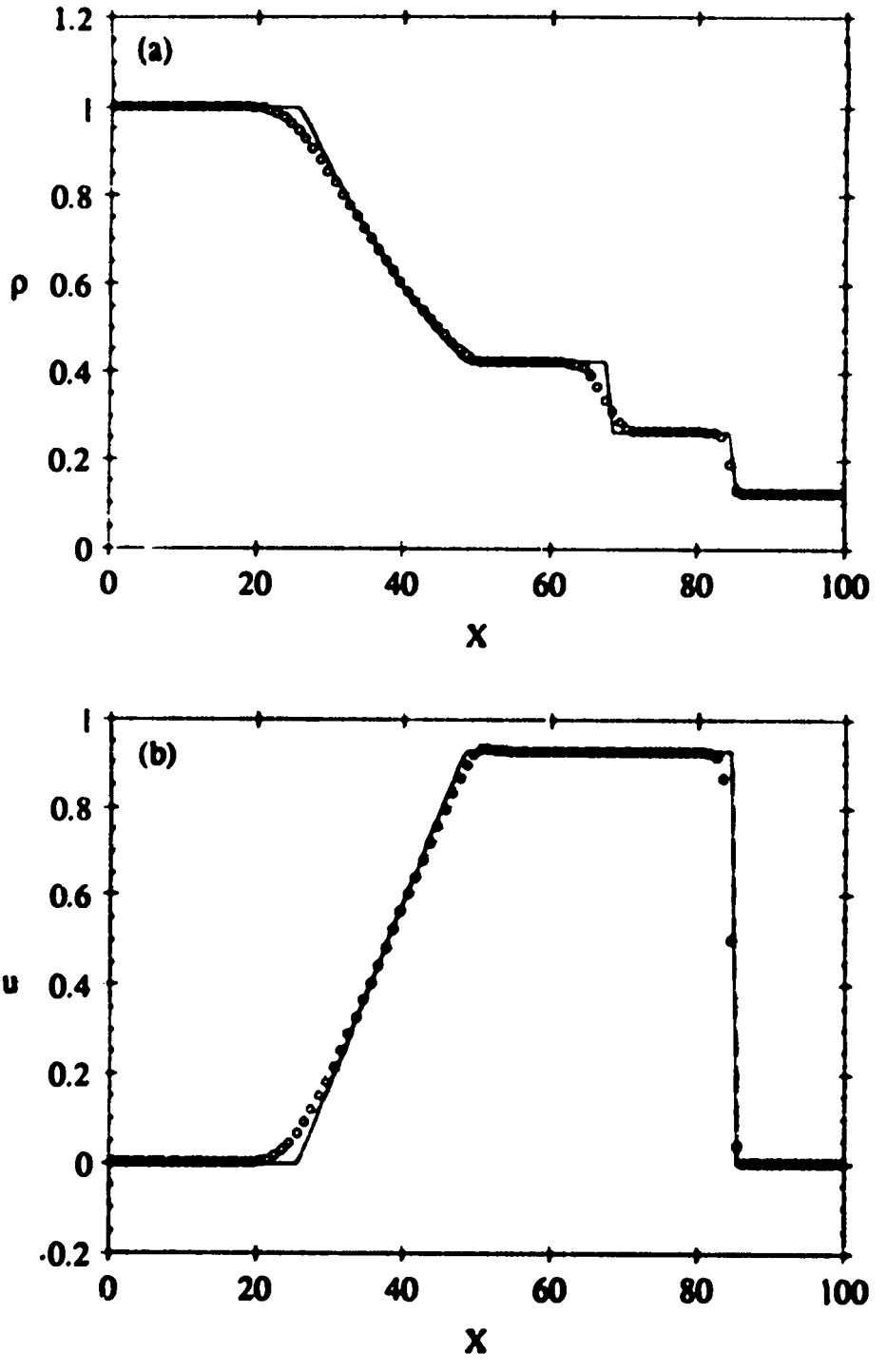


Figure 9.21: The density and velocity solutions to Sod's problem with an upwind-biased Lax-Wendroff TVD method with a cell-average correction.

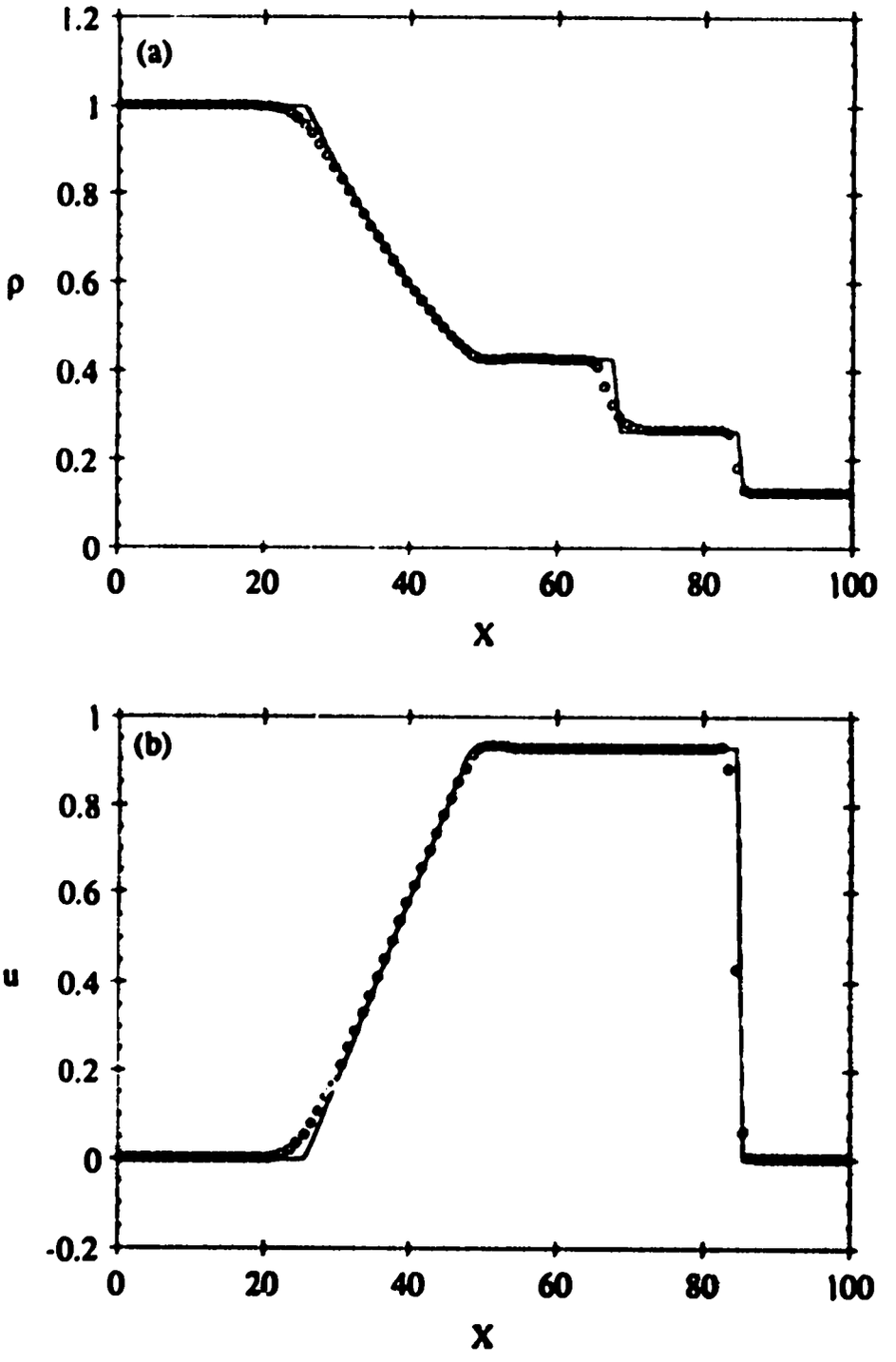


Figure 9.22: The density and velocity solutions to Sod's problem with a symmetric HOG method.

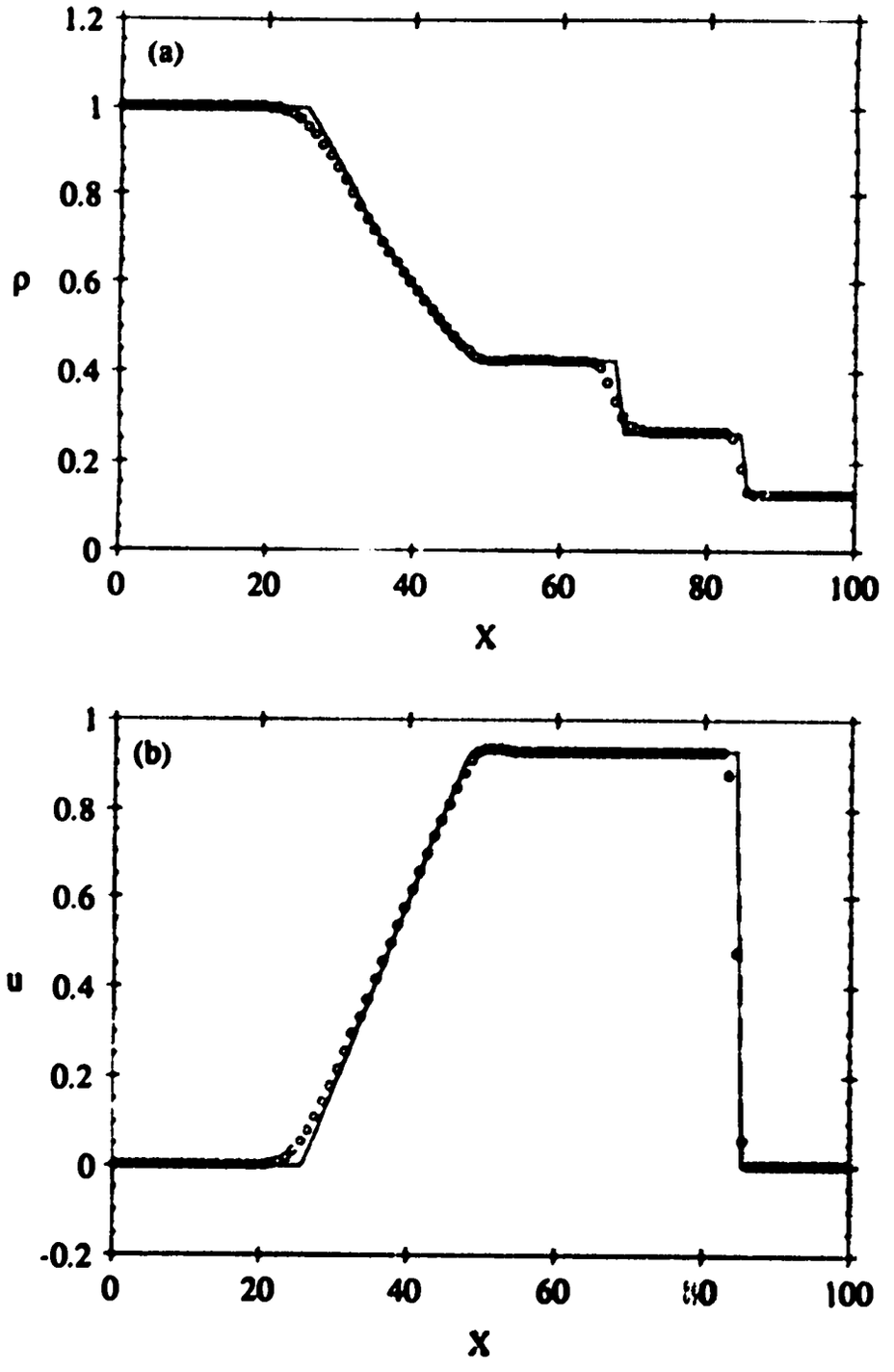


Figure 9.23: The density and velocity solutions to Sod's problem with a symmetric HOG method with a cell-average correction.

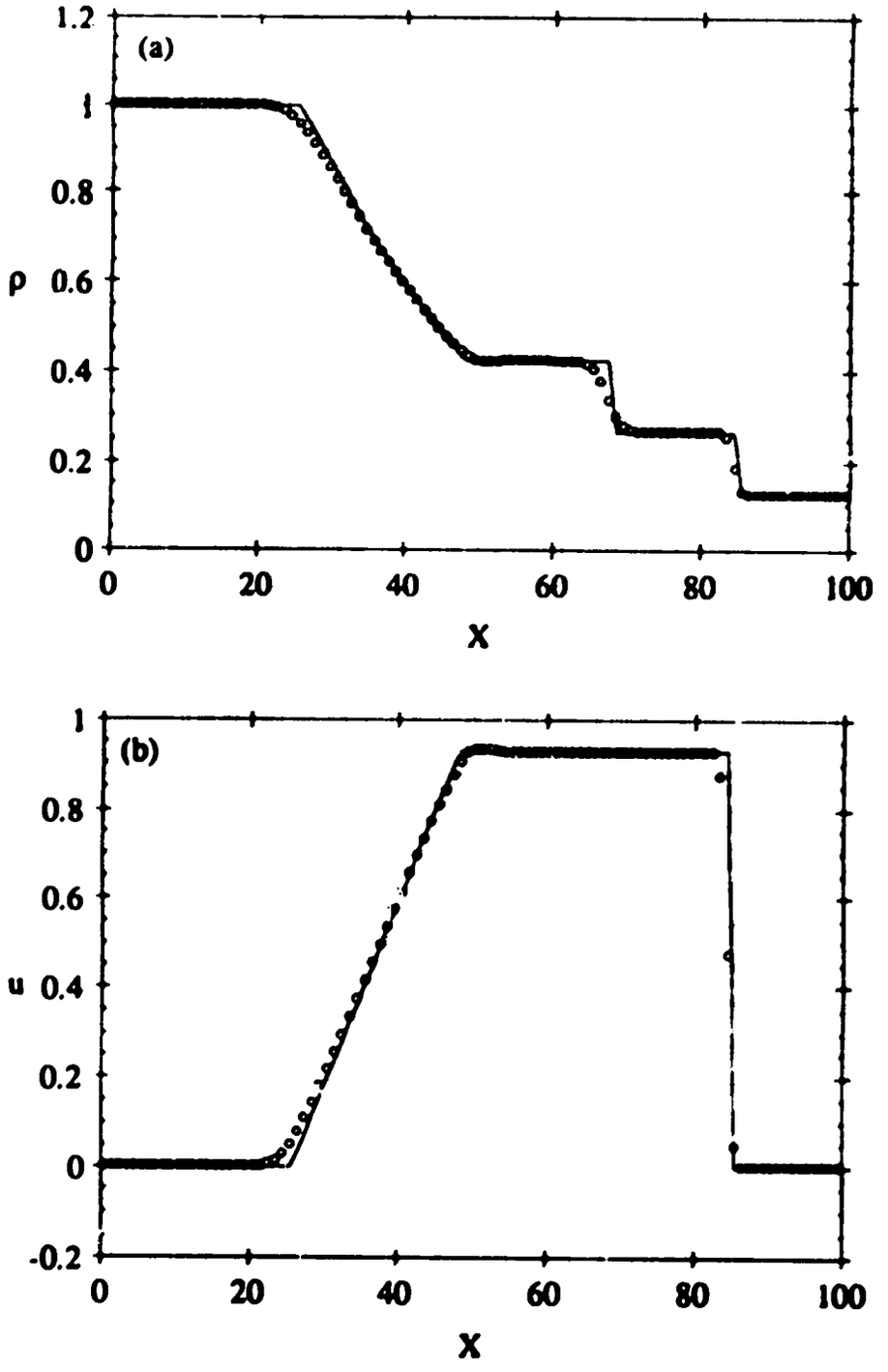


Figure 9.24: The density and velocity solutions to Sod's problem by a quadratic Taylor polynomial based HOG method.

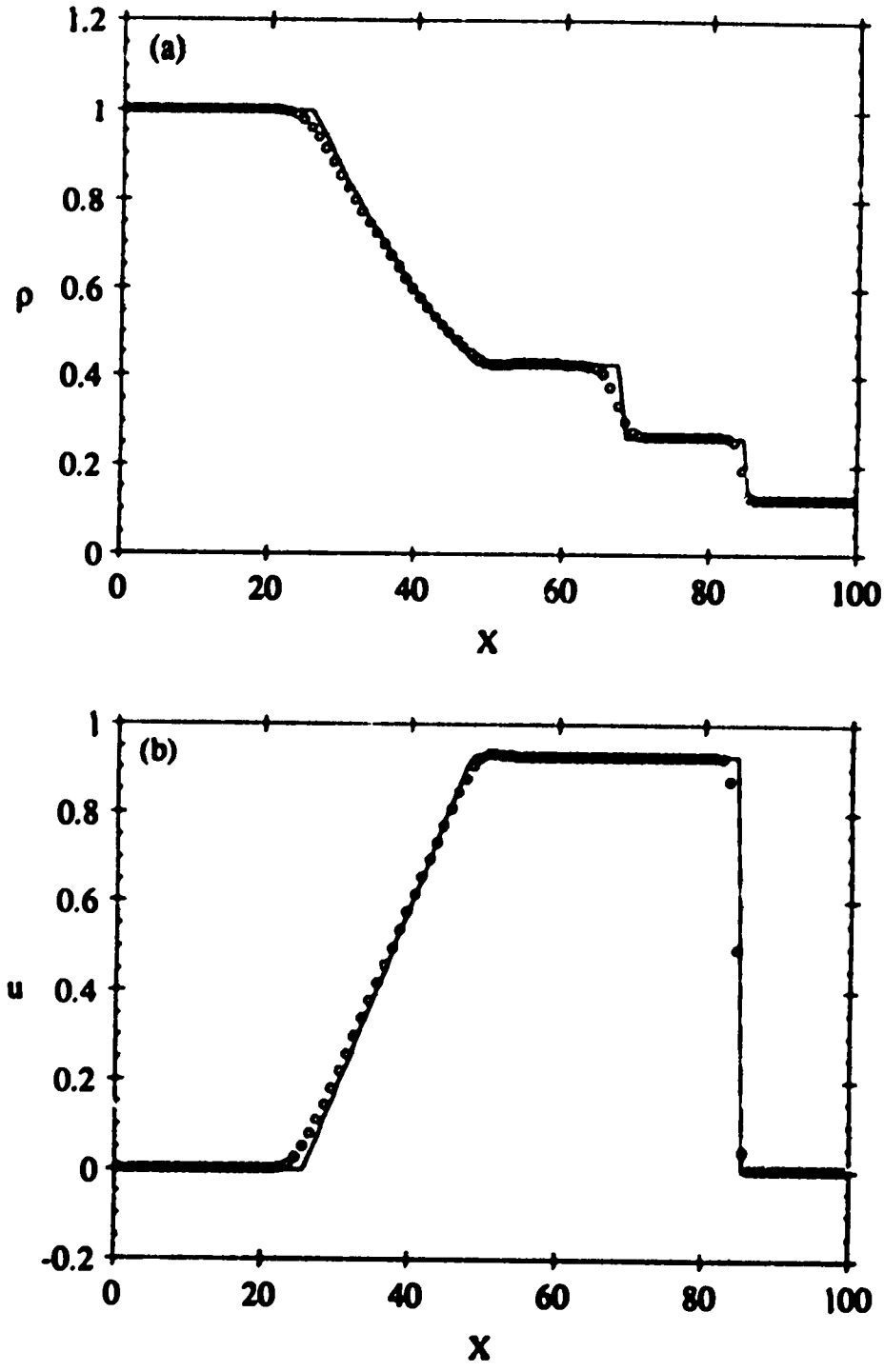


Figure 9.25: The density and velocity solutions to Sod's problem by a quadratic Legendre polynomial based HOG method.

9.5 Concluding Remarks

The results in the above section show that the method of reconstruction used in HOG schemes is of some importance to the quality of the results. For cases where the solution remain TVD, the cell-average solutions are of higher quality, but as the Burgers' equation solutions show, are of lower rates of convergence. Where the schemes are not TVD, the point-value reconstructions are superior and result in less oscillatory results. For systems of equations, the picture is less clear. The solutions obtained with all the methods show that the solutions are acceptable and quite good.

The major difference between the two approaches is one of ease of implementation. For one-dimensional problems, the differences are hardly consequential, but the edge is with the point-value polynomials. For multi-dimensional reconstructions, the point-value reconstruction is clearly easier and should be considered for this purpose despite certain philosophical inadequacies.

Chapter 10.
Conclusions and Recommendations

> Order and Simplification are the first steps toward the mastery of a subject. *Thomas Mann*
>
> Life is the art of drawing sufficient conclusions from insufficient premises. *Samuel Butler*

In this chapter, overall conclusions are made concerning the preceding work. These conclusions act as a summary of the results of this work. Following this a number of recommendations are made concerning future directions for research.

10.1 Conclusions

The FCT method is shown to be similar to symmetric TVD methods under certain conditions. This similarity is exploited in improving the performance of FCT. This improvement is particularly evident in the solution of systems of equations.

With the relationship between FCT and TVD methods firmly established, both of these methods were directly connected to high-order Godunov methods. This is accomplished through defining a non-upwind biased geometric version of the Lax-Wendroff method. Because the Lax-Wendroff method is the basis of the symmetric TVD method, the generalization is straightforward. From this, a scheme based on parabolic interpolation is derived. Further improvements are made through the use of uniformly non-oscillatory reconstruction methods.

The topic of limiters is then explored in considerable depth. This begins with a review of the FCT limiters. In this section of the work, Zalesak's limiter is modified in a similar fashion to the classic FCT limiter.

TVD limiters and their general properties are discussed in a manner that is more general than found in the literature. Three argument limiters are revised and extended with the use of certain limiter properties. The use of two parameter limiters is compared with three parameter limiters. It is shown that three parameter limiters induce a significant amount of numerical diffusion in a solution when compared to the analogous two parameter limiter. In addition, a general class of limiters referred to as nearly-TVD are discussed. These include TVB limiters, but also new classes of limiters such as generalized average limiters and S-Limiters. The ULTIMATE limiter is also discussed.

Finally, the topic of reconstruction in high-order Godunov methods is examined. This topic is precipitated by the work on high-order Godunov analogs to FCT/symmetric TVD methods. These high-order Godunov methods do not use a

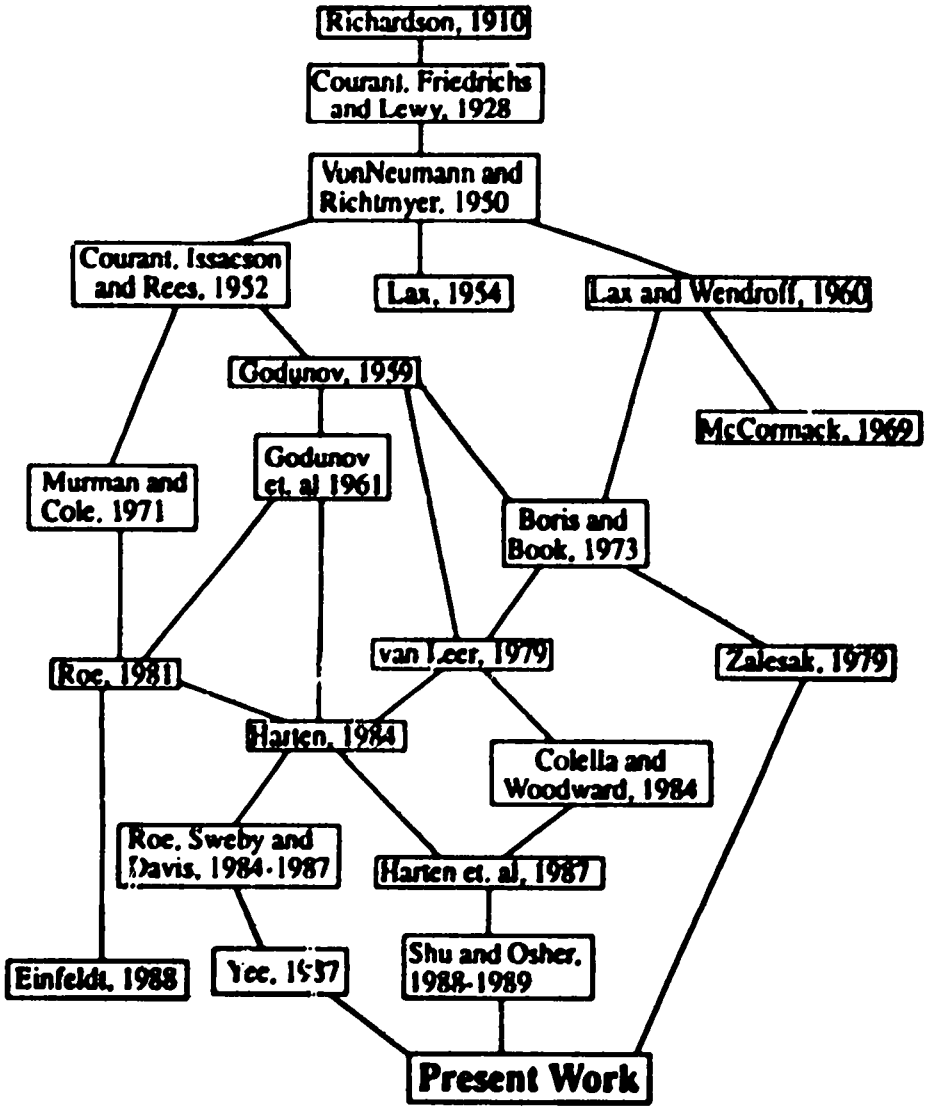


Figure 10.1: The significance of this work is shown in relation to the rough genealogy given in Chapter 2.

reconstruction step that requires the interpoiant to have an average value in a grid cell equal to the cell average in that grid cell. This property is discussed, and properties of the solutions using both standard and new high-order Godunov methods are examined. The lack of the cell-average property is demonstrated to not have significant negative consequences, and for certain situations to have positive consequences.

The principle advances made in this work can be seen graphically in Fig. 10.1. These conclusions can be summarized as follows:

- FCT was improved and shown to be part of a more general family of methods.
- Combined FCT and Symmetric TVD methods were extended into the HOC family of methods.

- A general procedure for improving FCT limiters has been described.
- A more general theory on limiters has been developed and used define new limiters.
- The difference between cell-average and point-value HOG schemes has been defined and explored. The point-value HOG schemes provide reliable solutions and improve on the cell-average HOG schemes when the scheme is not TVD.

10.2 Recommendations

With these conclusions in mind a number of recommendations for future research can be made. These do not cover the range of needed work, but represent some important needs from one perspective.

- In light of the results of this research and the literature, parabolic methods are worth exploring in much more detail. The added degree of freedom beyond linear interpolation allows the algorithm to be more flexible than second-order methods. Currently, the PPM method is the premier scheme for solving conservation laws. A large number of potential parabolic schemes exist, and should be studied in more detail. The use of parabolic schemes is need of assessment especially in the light of the results presented in Appendix F.
- One of the keys to the PPM algorithm is the use of a discontinuity detection algorithm [122]. This algorithm was the inspiration for the superbee limiter [132, 176]. The use of fuzzy logic [201, 202] should prove useful in designing this sort of algorithm. More generally, fuzzy limiters could have a more general application perhaps making limiters that work equally well in smooth and discontinuous regions of the flow.
- ENO methods should be broadened to include point-value schemes as well as the cell-average variety. In addition, other measures of reconstruction smoothness should be investigated perhaps using generalized average limiters is some sense. This is particularly important in the light of recent work [203].
- Smooth particle hydrodynamics (SPH) [204, 205, 206] may profit from nonlinear limiters. These methods typically use artificial viscosity to compute shocks. Through the use of biased gradient computations at discontinuities in the flow, (perhaps ENO-type algorithms) the use of artificial viscosity could be done away with. The resolution at these portions of the flow should also improve.
- Implicit numerical solutions with high resolution methods [196, 198, 207, 195, 15, 145] are important in aerospace applications. Currently artificial viscosity

methods [208, 209] are the preferred choice. The upwind type methods need to be more economical to compete. Research into multigrid acceleration of high-resolution upwind methods is a clear and present need. Also conjugate gradient type methods hold some promise [210]. The work of Yee and others [154] on nonlinear dynamics could provide some useful improvements.

- The role of Riemann solvers in algorithm dissipation is in need of clarification. Roberts [211] shows that the Riemann solver can cause oscillations for slow moving shocks even when used with Godunov's method. The solution is to use a more dissipative Riemann solver. This is important in light of the PPM's zone flattening algorithm, which is used to deal with such cases. This appears to be another place where fuzzy logic could be useful.

- The role of high-resolution upwind algorithms in turbulence research needs to be established. The work of Boris [77] is controversial with the large eddy simulation (LES) community. Others have used these methods in turbulence research with success [78, 212, 213, 79]. The results reported in [79] seem to show that high-resolution methods like the PPM give results indicative of very high Reynolds numbers. The impact of the design of methods on this use needs further assessment.

- Recently, front-tracking algorithms which are conservative have proven to be useful [129, 214, 215]. These coupled with adaptive mesh generation [129, 117] and high-order high-resolution methods are powerful solution methods. Coupling these methods to the design of new high resolution methods would be highly profitable. Other adaptive mesh algorithms [216, 217, 218] show promise. In addition techniques used in [219] may prove useful.

- The use of these methods in radiation transport may be applicable. In discrete ordinates methods [220, 221] diamond differencing is typically used, although linear discontinuous methods also are used. Both of these methods could profit from modern upwind methods to insure positivity of solutions. The linear discontinuous method has been used for high resolution fluid flow solutions [222].

- Multiphase flow presents a number of challenges to the use of this sort of method. Typically, the algorithms used for this type of flow are semi-implicit [3, 2, 223]. Semi-implicit time discretizations are in need of development and would be useful in other applications [93] where problems are stiff in some manner. Multiphase flow can also be ill-posed in the sense of Haramard, thus creating difficulty with Riemann solvers.

- Multidimensional schemes are an active topic of research. The role of limiters and their form is an open question. The methods of Akima [224, 225, 226, 227]

may prove useful in defining multidimensional limiters. The use of ENO schemes in multidimensions is proceeding [139, 199, 138], but it is in its infancy.

- Multidimensional Riemann solvers need work. Most current schemes show poor results because they are not monotone (based on a wave analogy [228]). Recent work on flux-splitting in several dimensions [229] may prove very useful in a number of regards and needs further development.

Other research is also exciting. The use of high-resolution upwind methods with incompressible flow computations, weather simulations and other applications [188, 230] shows considerable promise.

Appendix A.

Test Problems

A.1 Introduction

The methods described in this research are used to solve three test problems: the scalar wave equation

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 , \tag{A.1a}$$

inviscid Burgers' equation

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right) = 0 . \tag{A.1b}$$

and the Euler equations (see Appendix B for a more complete discussion) for an ideal gas

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0 , \tag{A.1c}$$

where

$$\mathbf{U} = \begin{bmatrix} \rho \\ m \\ E \end{bmatrix} , \mathbf{F} = \begin{bmatrix} m \\ m^2/\rho + p \\ m(E+p)/\rho \end{bmatrix} .$$

For the Euler equations the variables are defined $m = \rho u$ where u is the fluid velocity, density, ρ, and the pressure, p, are related to the energy, E, by an equation of state *(for an ideal gas)*,

$$p = \rho\varepsilon(\gamma - 1) ,$$

where $\varepsilon = E/\rho - 1/2u^2$ and γ is the ratio of specific heats for the gas in question.

A.2 Scalar Wave Equation

In this section, the test problems used for the scalar wave equation are described. Four initial conditions are used for the analysis: a square wave with a width of 10 cells, a sine wave over one full period with a width of 20 cells, a sine squared wave (half of a period) of a width of 25 cells and a triangle function with a width of 10. The advective velocity is taken to be unity. Each of these test problems is shown in Fig. A.1. The exact solution for the scalar wave equation is given by

$$u(x,t) = u_0(x - at) . \tag{A.2}$$

where a is the advective velocity and $u_o(x)$ is the initial condition.

The course appearance of several of the figures is misleading. The two functions based on $\sin(x)$ are smooth. The course nature of the plots results from the low resolution of the discretization.

A.3 Burgers' Equation

The test problem consists of N equidistantly spaced cells on a domain $x \in [0, 2\pi]$. The initial condition is $\sin(x)$. At $t = 0.2$ and $t = 1.0$ the solution is compared with the exact solution. At $t = 0.2$ the solution is smooth; however, at $t = 1.0$ the solution has developed a shock. The CFL number is ≈ 0.4. The solutions at these two times are shown in Fig. A.2. The exact solution is produced using a formula found in [67], which is

$$u(x,t) = \frac{\partial}{\partial x} \min_y \left[\int_o^y u_o(x)\,dx + \frac{1}{2t}(x-y)^2 \right] , \tag{A.3}$$

where the definitions are the same as for the scalar wave equation.

A.4 The Euler Equations

The Euler equations are used as an example of the solution process on a system of equations. The Euler equations are perhaps the most common application of the methods discussed in this work.

A.4.1 Sod's Problem

The problem used by Sod [41] to test a number of methods for solving the equations of compressible flow has become a standard test problem. The initial condition for this problem consists of two semi-infinite states separated at $t = 0$, and the left and right states are set to the following conditions:
for $X < 50.0$,

$$\begin{bmatrix} \tau_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 1.0 \end{bmatrix} ,$$

and for $X \geq 50.0$

$$\begin{bmatrix} \tau_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 8.0 \\ 0.0 \\ 0.1 \end{bmatrix} .$$

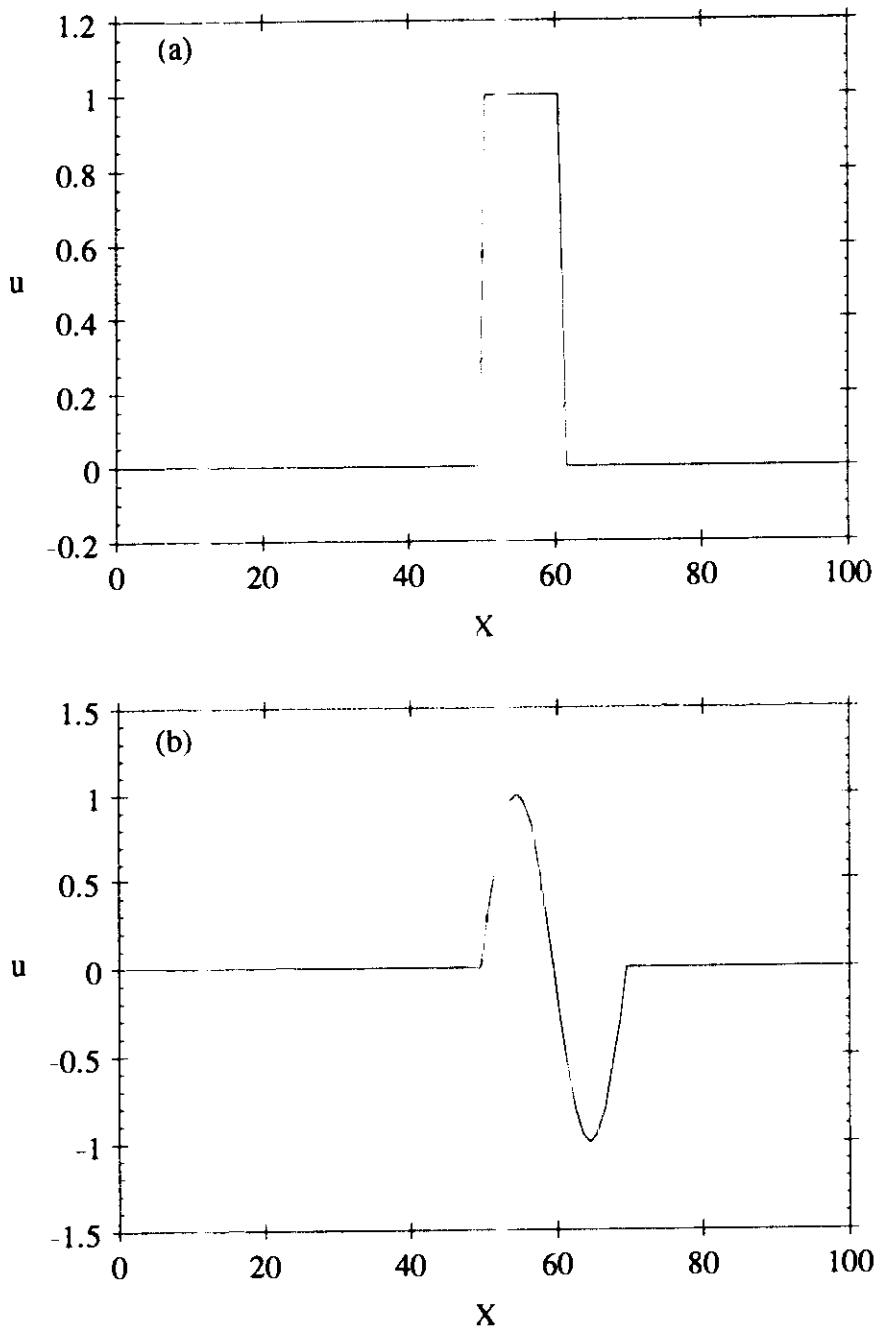


Figure A.1: The exact solutions to the test problems used in the scalar wave equation tests. These are the square wave, sine wave, sine squared wave and the triangle wave.

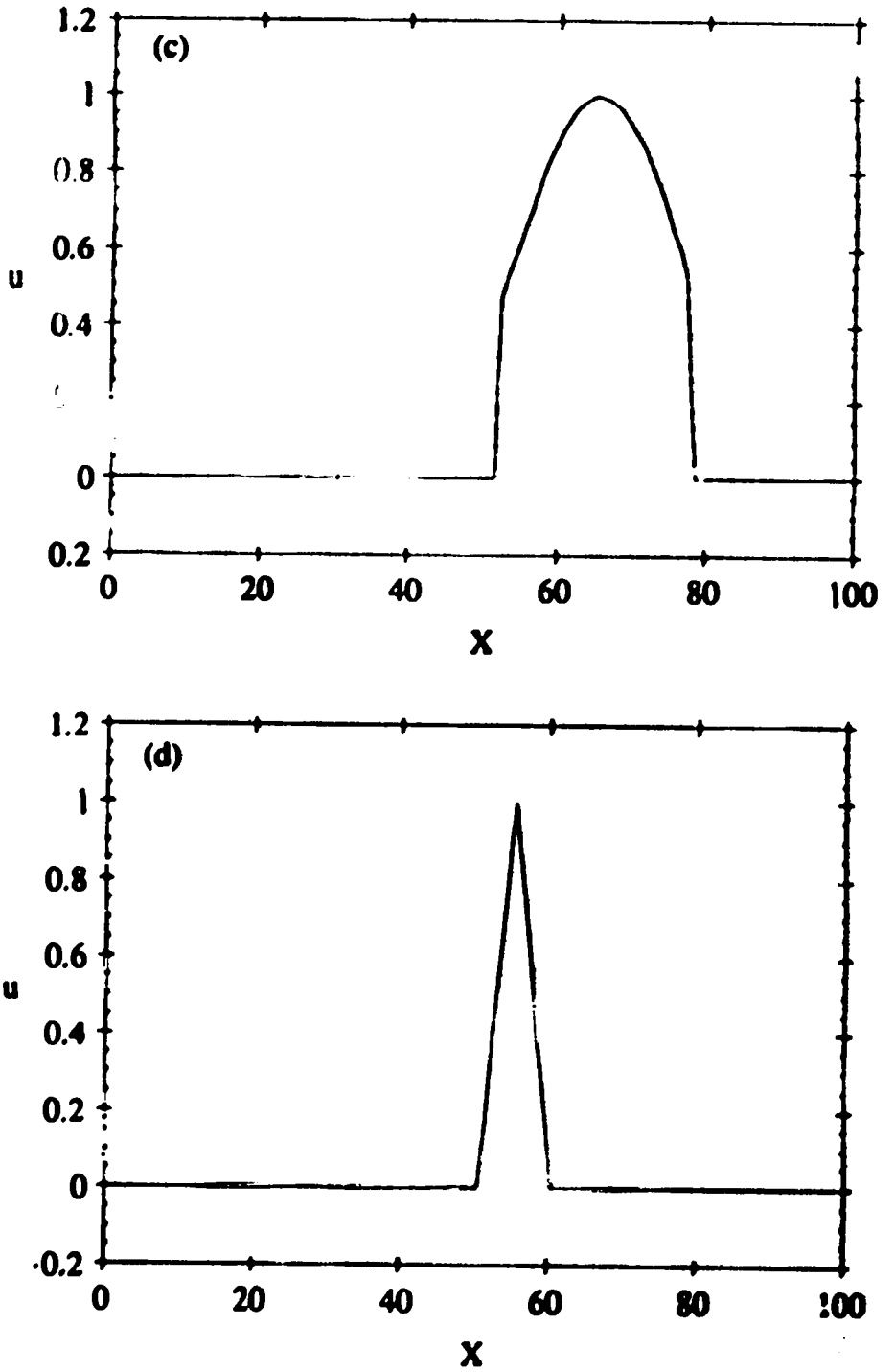


Figure A.1: continued.

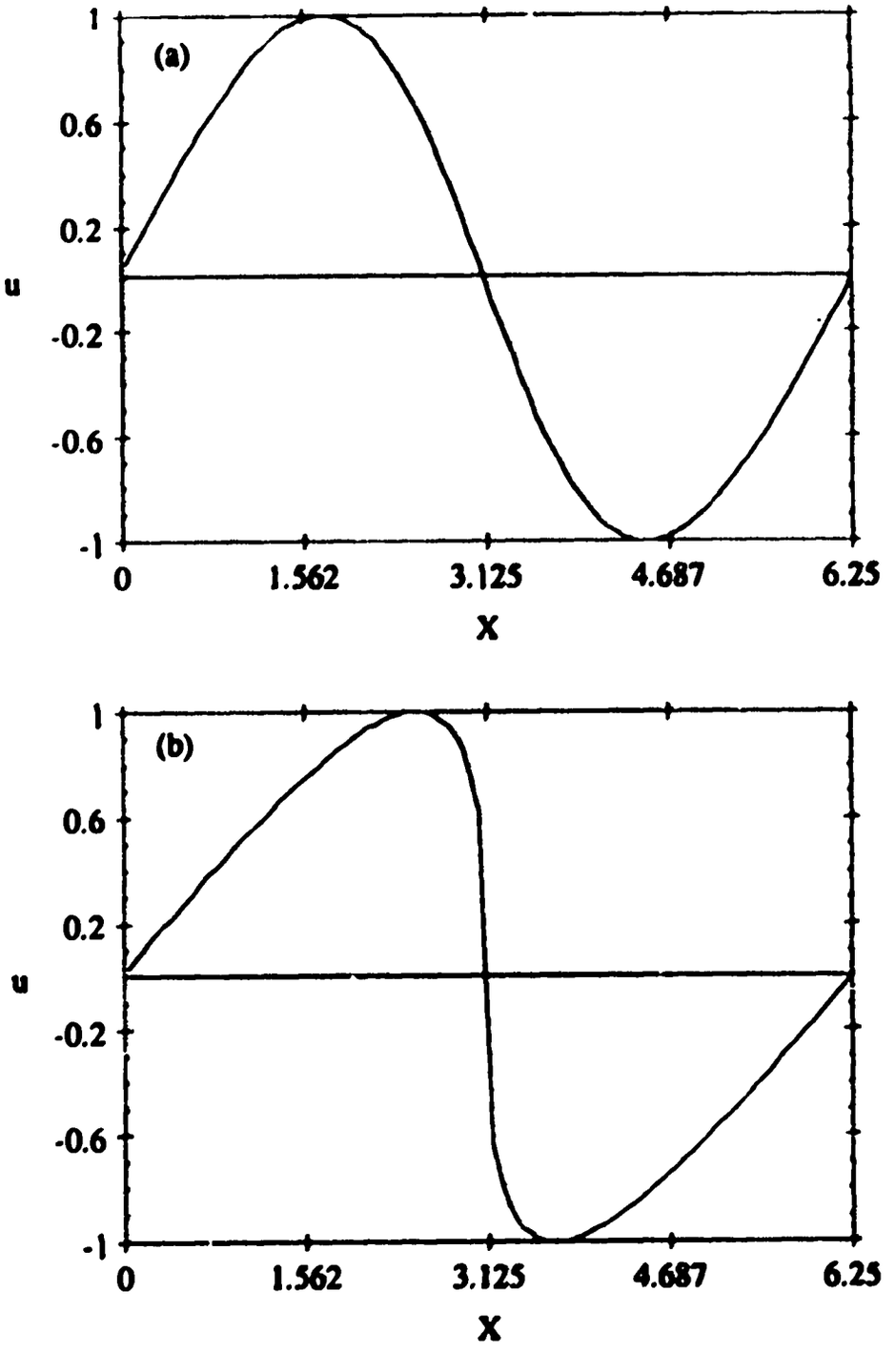


Figure A.2: The exact solutions to the test problems used in the Burgers' equation tests. The figures are shown at $t = 0.2$ in (a) and $t = 1.0$ in (b).

with $\gamma = 1.4$. The domain is discretized into 100 cells of equal lengths ($\Delta x = 1.0$) and the CFL number is set to 0.9. The solutions are shown in Fig. A.3 at $t = 20$. The exact solutions can be seen in Fig. A.3. These solutions are computed with the method described in Appendix B for the exact solution to a shock tube problem.

A.4.2 Lax's Problem

Lax's problem is a shock tube problem similar to Sod's, but with one of the two semi-infinite states used as initial conditions not being at rest. The initial condition for this problem consists of two semi-infinite states separated at $t = 0$, the left and right states are set to the following conditions:
for $X < 50.0$,

$$\begin{bmatrix} \tau_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 2.24 \\ 0.69 \\ 3.52 \end{bmatrix},$$

and for $X \geq 50.0$,

$$\begin{bmatrix} \tau_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 2.0 \\ 0.0 \\ 0.57 \end{bmatrix},$$

with $\gamma = 1.4$. The domain is discretized into 100 cells of equal lengths ($\Delta x = 1.0$) and the CFL number is set to 0.9. The solutions are shown in Fig. A.4 at $t = 15$. The exact solution can be seen in Fig. A.4.

A.5 The Vacuum Problem

The vacuum problem is a shock tube problem where two identical states are moving away from each other at $t = 0$. The states are kinetic energy rich, which causes problems for the finite difference schemes. The initial condition for this problem consists of two semi-infinite states separated at $t = 0$, the left and right states are set to the following conditions
for $X < 50.0$,

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ -2.0 \\ 1.0 \end{bmatrix},$$

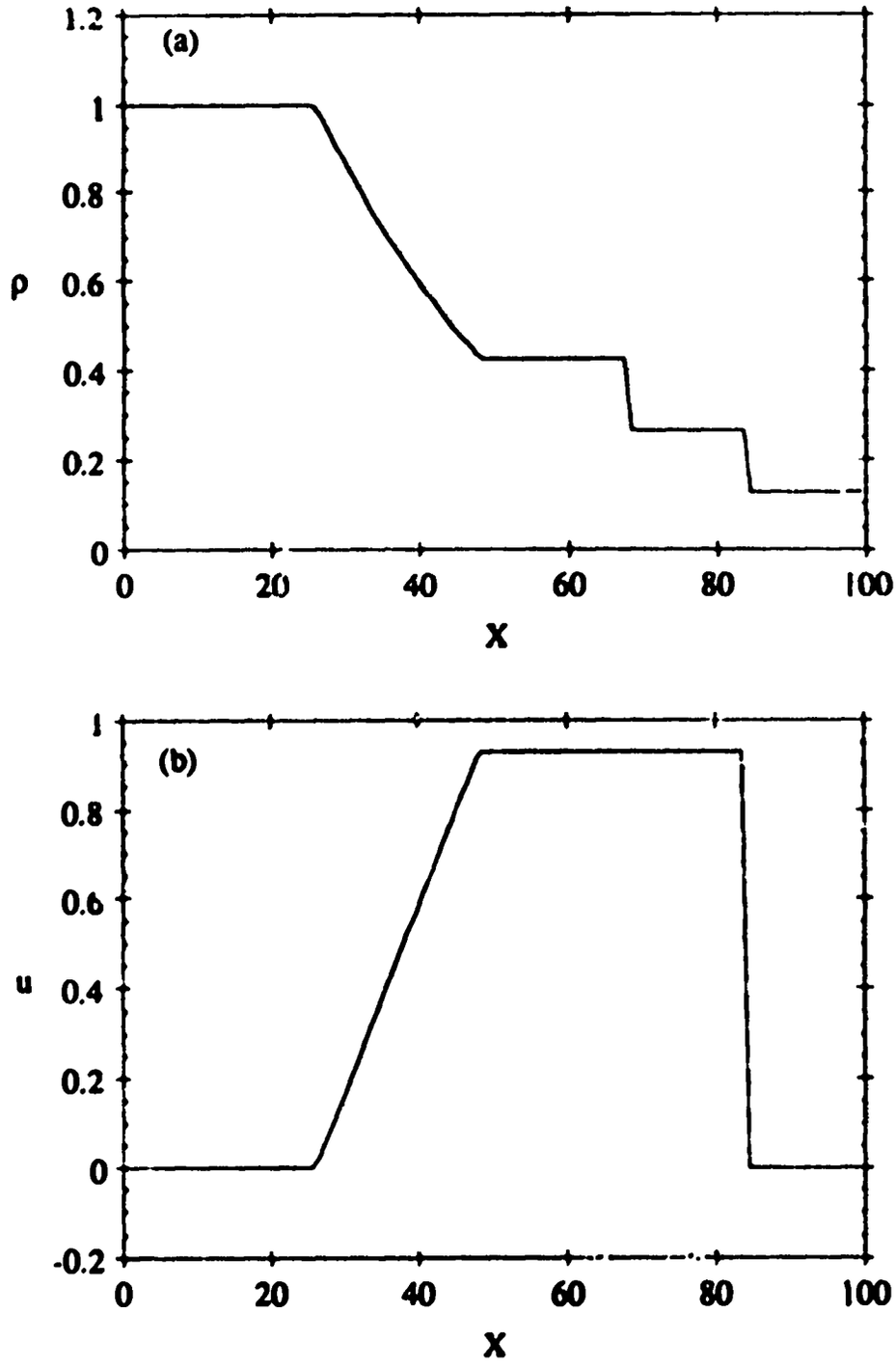
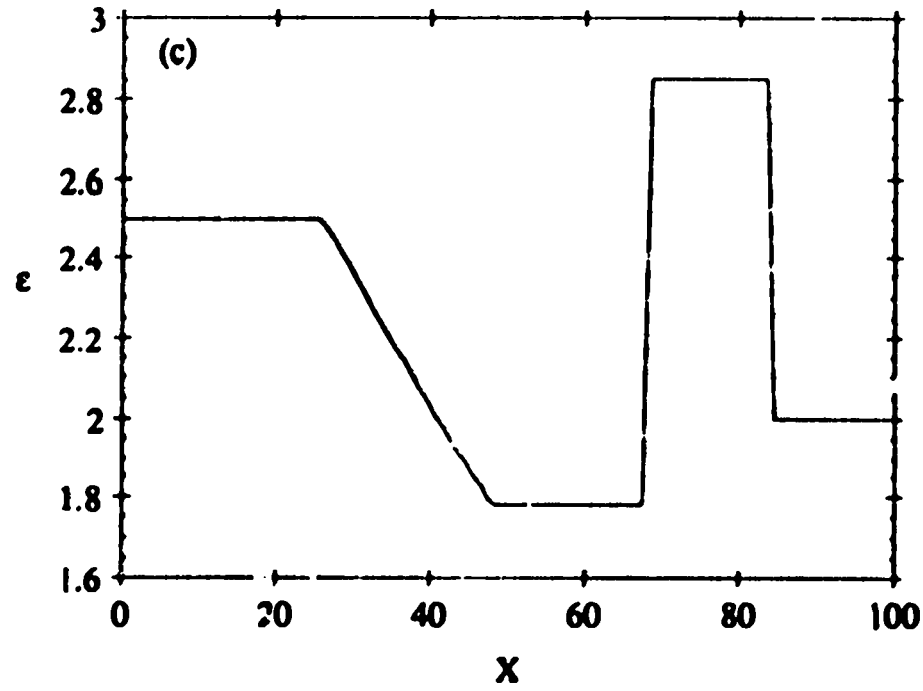


Figure A.3: The exact solution for Sod's Riemann problem. Note the appearance of the rarefaction wave running from about $x \approx 30$ to $x \approx 50$, which is a smooth transition. The contact discontinuity is at about $x \approx 65$ and the shock is at $x \approx 85$. Note that the transitions between states for these two structures are sharp. The density and energy profiles show more structure than the velocity or pressure profiles because of the contact discontinuity.



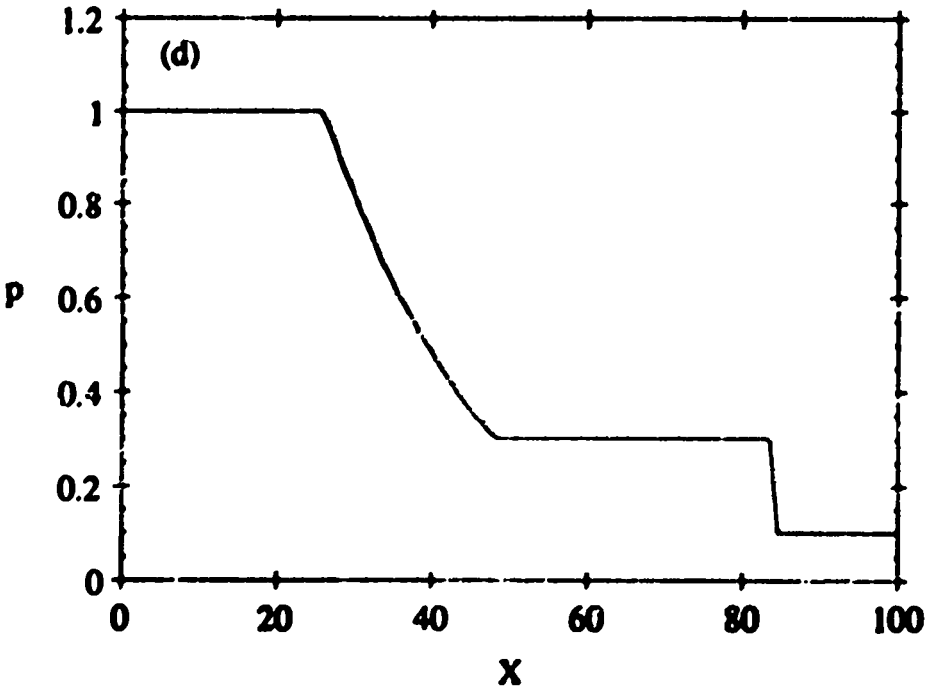


Figure A.3: continued.

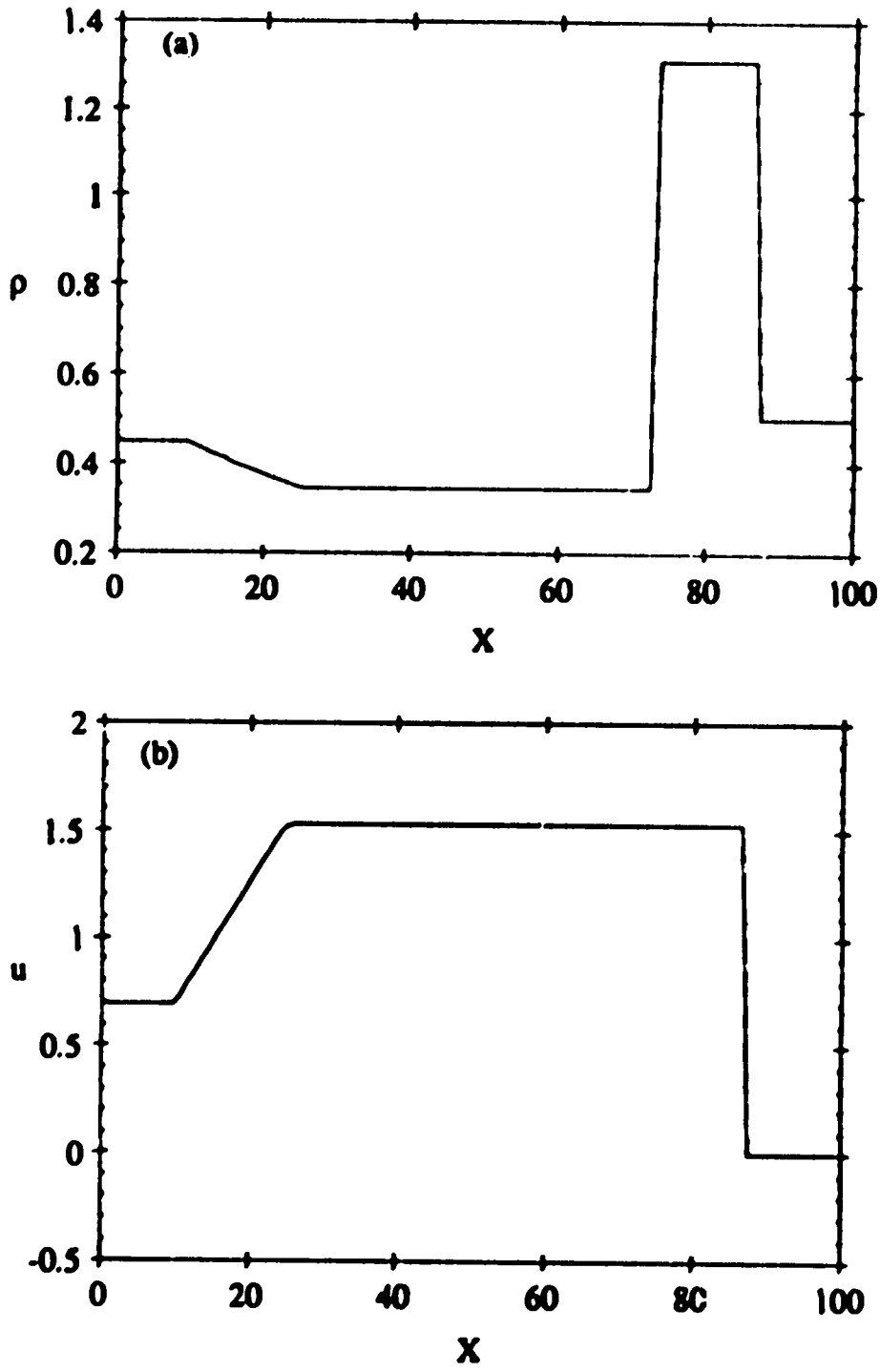


Figure A.4: The exact solution for Lax's Riemann problem. Note the appearance of the rarefaction wave running from about $x \approx 10$ to $x \approx 25$, which is a smooth transition. The contact discontinuity is at about $x \approx 75$ and the shock is at $x \approx 90$.

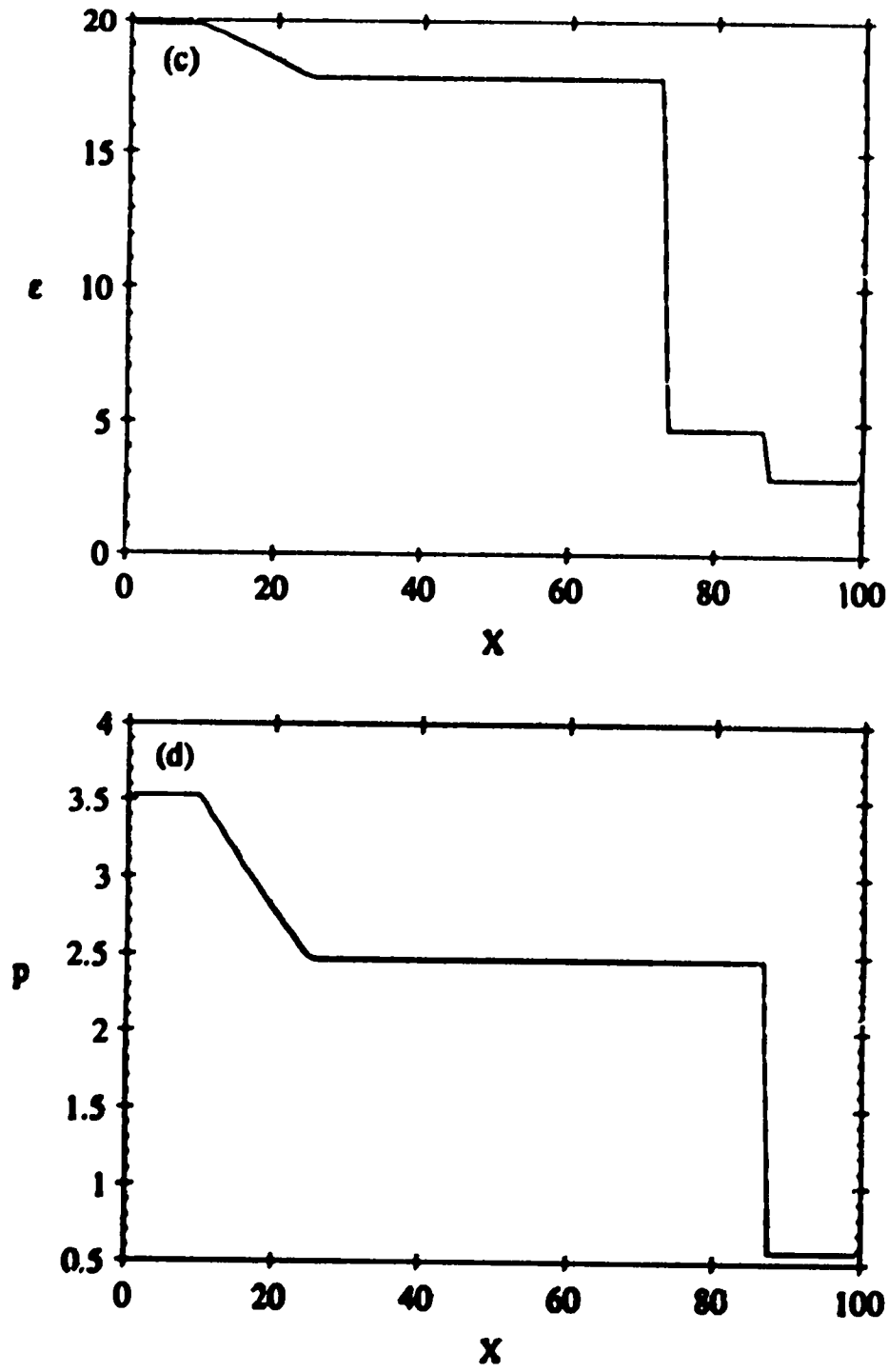


Figure A.4: continued.

and for $X \geq 50.0$,

$$\begin{bmatrix} \rho_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 1.0 \\ 2.0 \\ 1.0 \end{bmatrix},$$

with $\gamma = 1.4$. The domain is discretized into 100 cells of equal lengths ($\Delta x = 1.0$) and the CFL number is set to 0.9. The solutions are shown at $t = 10$. An additional caveat is that the computation of the stability criteria also involves the condition based on a condition similar to the "tangling" or "emptying" conditions in Lagrangian computations, i.e.,

$$\Delta t \leq C \frac{\Delta x}{\left| u_{j+\frac{1}{2}} - u_{j-\frac{1}{2}} \right|}, \tag{A.4}$$

where $C \in [0, 1]$. The exact solution can be seen in Fig. A.5.

A.5.1 Blast Wave Problem

This blast wave problem was used by Woodward and Colella [44] to test a variety of high-resolution methods. This test turns out to be an extremely stringent test of numerical methods for solving hyperbolic conservation laws. The initial conditions consist of the following:
for $X \leq 10.0$,

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 1000.0 \end{bmatrix},$$

for $10.0 > X > 90.0$,

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 0.01 \end{bmatrix},$$

and for $X \geq 90.0$

$$\begin{bmatrix} \rho_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 100.0 \end{bmatrix},$$

with $\gamma = 1.4$. The boundary conditions play an important role in this problem and are reflective at both the left ($X = 0$) and right ($X = 100$) walls. The solutions are shown in Fig. A.6 at $t = 3.50$. The solution develops into two strong shock waves that

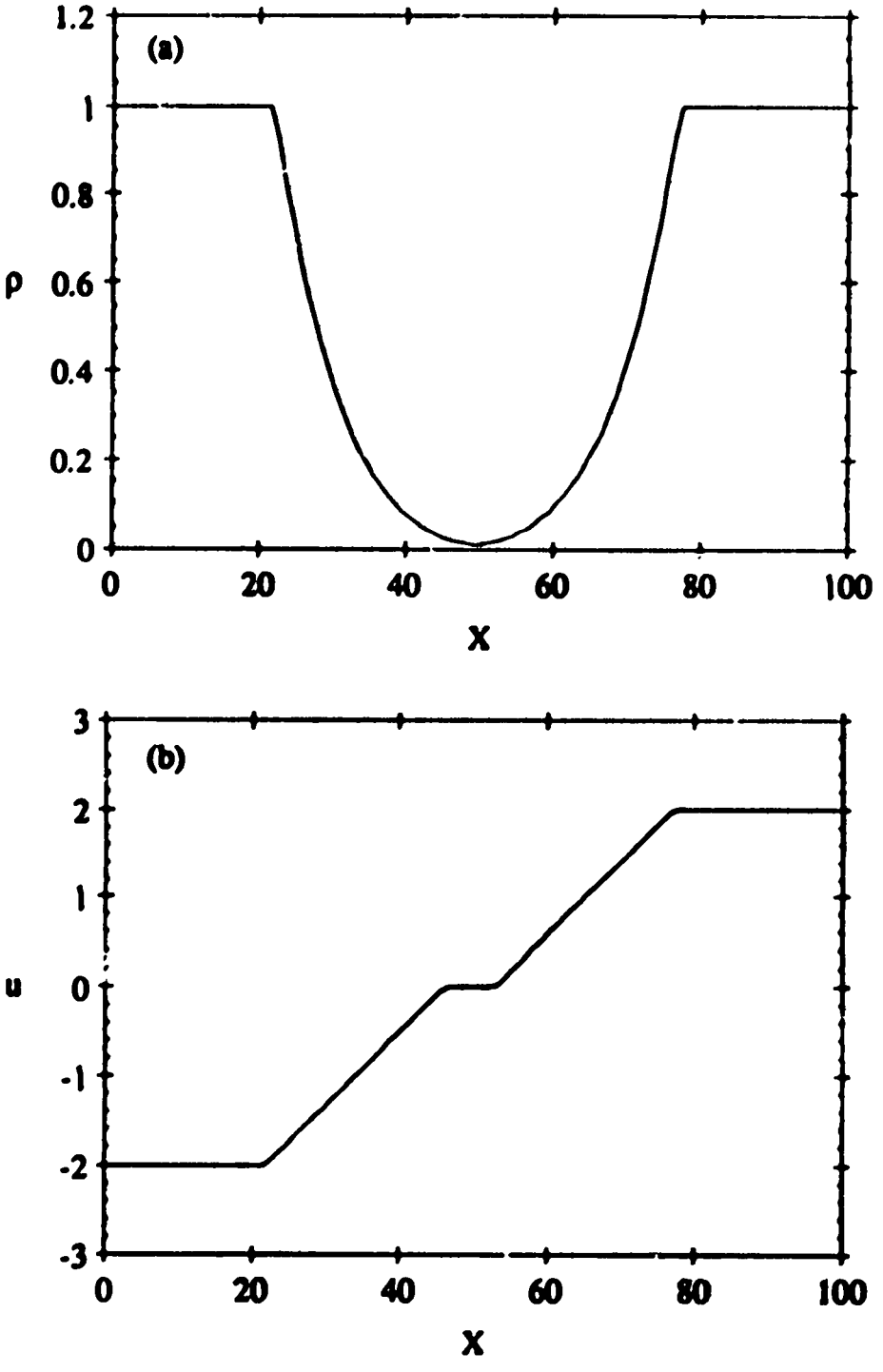
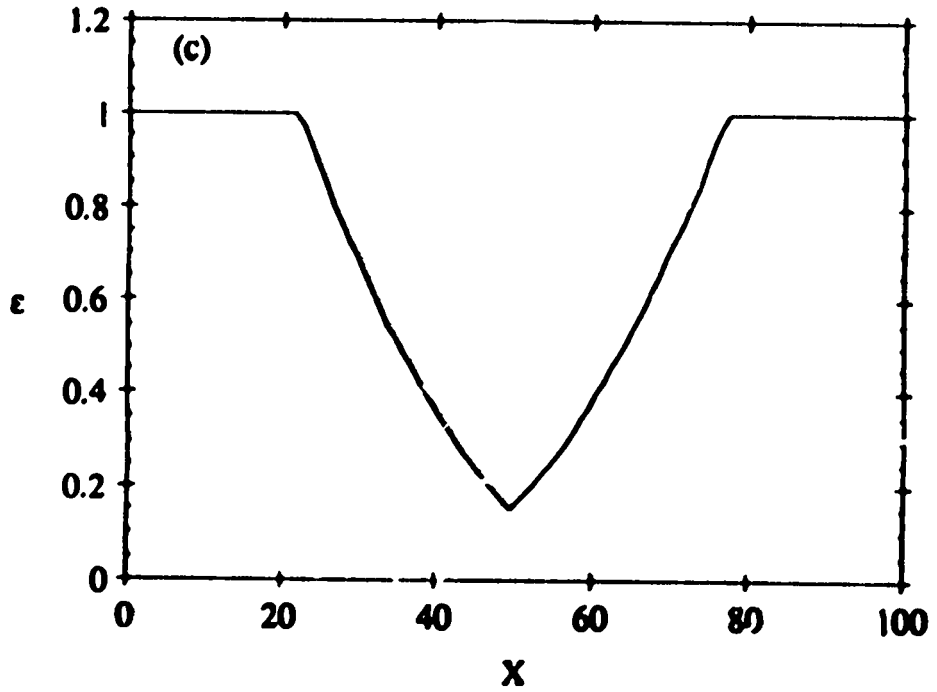


Figure A.5: The exact solution for the vacuum Riemann problem. Note the appearance of the rarefaction waves running both directions from the initial discontinuity. The internal energy plot (c) shows error near the vacuum because of round off errors.



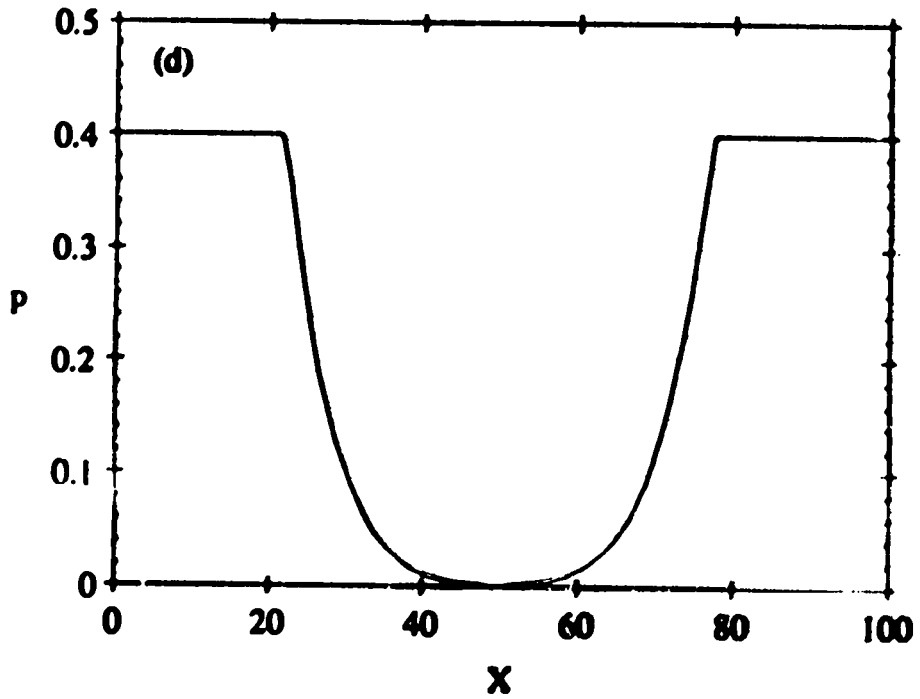


Figure A.5: continued

collide. The result of this is a complex set of shock and rarefaction waves as well as contact discontinuities in a small region of space. These interactions are exceedingly difficult to resolve on a fixed Eulerian grid without prior knowledge of the solution so that the grid can be locally refined (certain adaptive meshing procedures can avoid the need for a priori knowledge of the solution). The "exact" solution can be seen in Fig. A.6. This "exact" solution was computed with 2000 grid cells at a CFL number of 0.95 using a cell-centered second-order HOG method. The superbee limiter was used on the linearly degenerate field and van Leer's limiter was used on the nonlinear fields (see Chapter 8 for a complete discussion of the limiters).

The solution of Riemann problems both exactly and approximately is discussed in the next appendix.

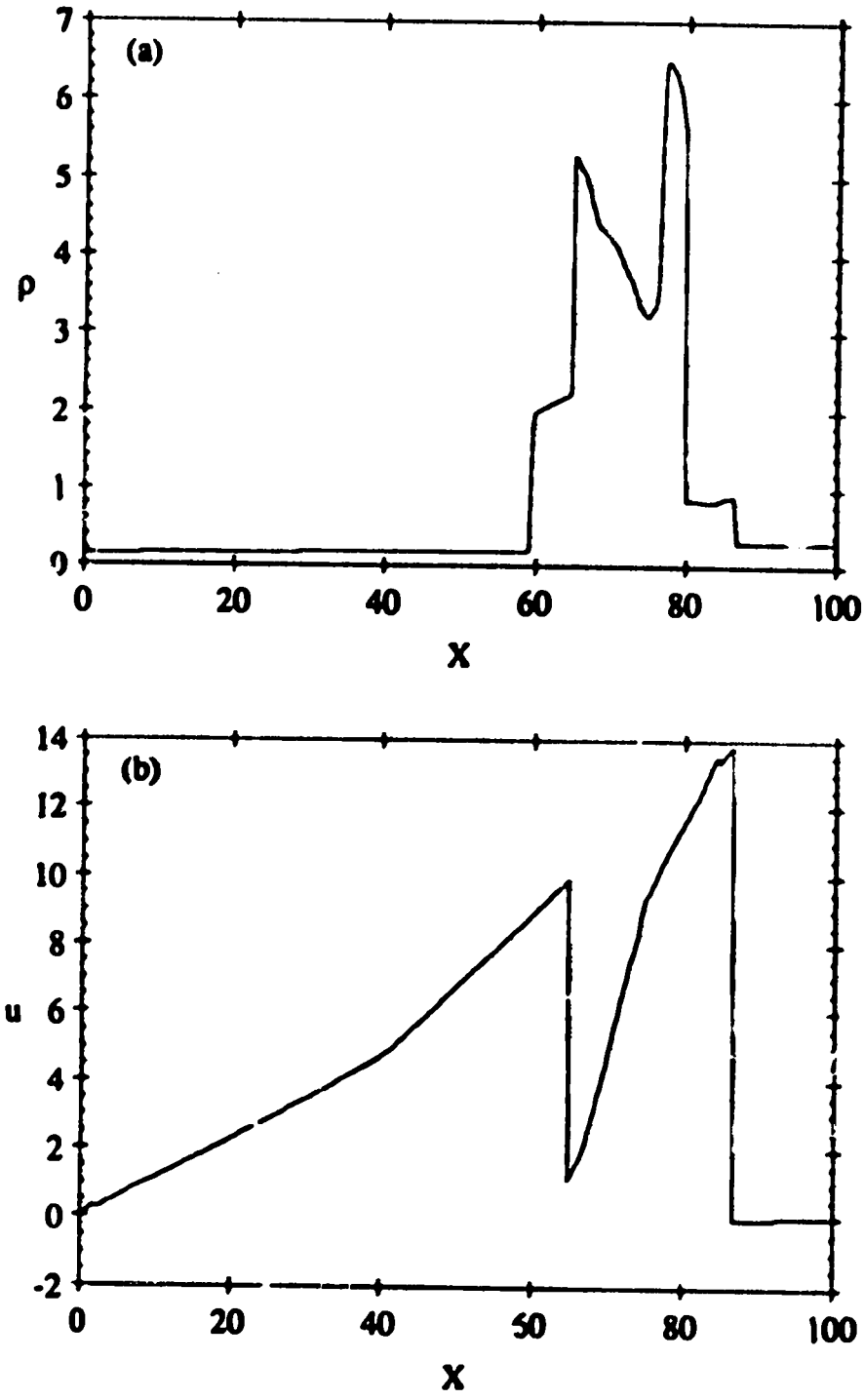


Figure A.6: The "exact" solution for the blast wave problem. Note the large amount of solution structure between $x \approx 60$ and $x \approx 85$. The two strong blast waves are interacting and are in the process of passing through one another. The interaction region is richly populated with contact discontinuities and shock waves.

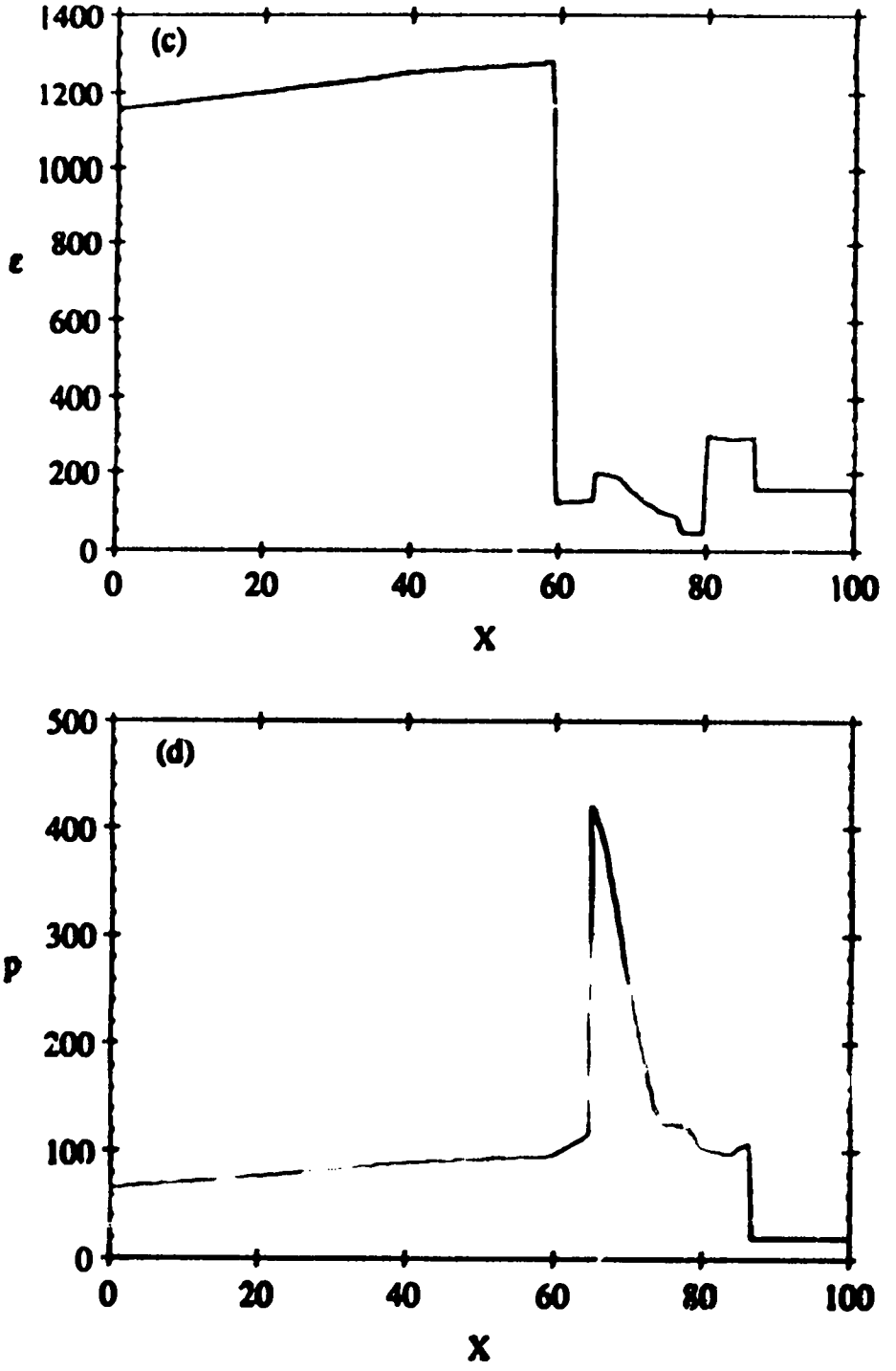


Figure A.6: continued.

Appendix B.

The Equations of Compressible Flow and Riemann Solvers

B.1 Introduction

When developing solution techniques for the equations of compressible flow, the common practice is to solve the equation in a Eulerian frame of reference. For certain classes of problems, a Lagrangian or Lagrangian followed by a transiation back to an Eulerian frame methods has advantages.

Much of the development of current high-resolution numerical methods for the solution of the Euler equations was the product of just such algorithms, although development has concentrated on purely Eulerian schemes in recent years. Godunov's method [56] is the basis of van Leer's work [60]. These methods find the solution to a Lagrangian flow system and then remaps it to a fixed (or moving) Eulerian grid. This methodology can also be thought of as operator splitting [156] based on convective and sound waves. The piecewise parabolic method [122] extended van Leer's method. Godunov and coworkers also introduced a purely Eulerian variant of Godunov's method [57], which can be thought of as the basis of current purely Eulerian methods.

In modern high-resolution Eulerian algorithms, it is common to use approximate Riemann solvers of some sort to compute correct wave propagation, because exact Riemann solvers [60, 41] are expensive. As a solution to this problem, several researchers have developed approximate Riemann solvers. Each of these has seen its primary development and use in an Eulerian frame. In this appendix, seven types are explored:

1. a naive Riemann solver,
2. the Lax-Friedrichs Riemann solver [55],
3. the local Lax-Friedrichs (LLF) Riemann solver [65, 66],
4. the simple Riemann solver introduced in [30] and refined in [126, 231], known as the HLLE (Harten, Lax, van Leer and Einfeldt) Riemann solver,
5. Roe's approximate Riemann solver [63], discussed in [232],
6. the Riemann solver of Engquist and Osher [127],

7. and flux splitting like that of Steger and Warming [125].

The next section discusses the flow solution algorithms and derives several approximate Riemann solvers.

B.2 The Equations of Compressible Flow

The Euler equations represent the conservation of mass, momentum, and energy in a fixed coordinate system and in one dimension are

$$\frac{\partial \rho}{\partial t} + \frac{\partial m}{\partial x} = 0 , \tag{B.1a}$$

$$\frac{\partial m}{\partial t} + \frac{\partial}{\partial x}\left(\frac{m^2}{\rho} + p\right) = 0 . \tag{B.1b}$$

and

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x}\left(\frac{m}{\rho}(E + p)\right) = 0 . \tag{B.1c}$$

Here ρ is the density, m is the momentum ($m = \rho u$, where u is the flow velocity), and E is the total energy. The other variables are related to the pressure p through an equation of state

$$p = f(\rho, i) , \tag{B.1d}$$

where $i = E/\rho - \frac{1}{2}u^2$. For an ideal gas, the equation of state is $p = \left(E - \frac{1}{2}m^2/\rho\right)(\gamma - 1)$ with γ being the ratio of specific heats. This system is hyperbolic and has three characteristic velocities $u - c$, u, and $u + c$, where c is the sound speed. For an ideal gas

$$c^2 = \frac{\gamma P}{\rho} .$$

Fairly directly, this system can be converted to a system of equations in conservation form for a coordinate system moving at the flow velocity. This introduces a change of coordinates from the variable x to ξ where ξ is the mass coordinate defined by

$$\xi = \int \rho dx , \text{ or } d\xi = \rho dx .$$

The system of equations is then

$$\frac{\partial \tau}{\partial t} - \frac{\partial u}{\partial \xi} = 0 , \tag{B.2a}$$

$$\frac{\partial u}{\partial t} + \frac{\partial p}{\partial \xi} = 0 , \tag{B.2b}$$

and

$$\frac{\partial e}{\partial t} + \frac{\partial pu}{\partial \xi} = 0 . \tag{B.2c}$$

In this equation set $\tau = 1/\rho$ and $e = \tau E$. This system also has three characteristic speeds: $-C$, 0, and C, where $C^2 = \gamma p/\tau$ is the Lagrangian sound speed for an ideal gas. The ideal gas equation of state in terms of the Lagrangian variables is $p = \left(e - \frac{1}{2}u^2\right)(\gamma - 1)/\tau$.

With remap equations the solutions found with these equations can be remapped to an Eulerian grid (as is discussed in the next section) and produce a solution that is equivalent to the solution of the first equation set. The three remap equations are

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} = 0 , \tag{B.3a}$$

$$\frac{\partial m}{\partial t} + u\frac{\partial m}{\partial x} = 0 , \tag{B.3b}$$

and

$$\frac{\partial E}{\partial t} + u\frac{\partial E}{\partial x} = 0 . \tag{B.3c}$$

B.3 Solution Algorithms

In this section, Godunov's method is described with specific attention being given to the Lagrangian formulation with an Eulerian remap. This is followed by a discussion of each of the approximate Riemann solvers used in this study.

B.3.1 Exact Solution of the Riemann Problem

The construction of the exact solution to the given Riemann problem follows the algorithm given in Sod's paper [41] with improvements suggested in [60, 177]. These improvements constitute a Newton-type iteration to solve the nonlinear governing equations for the Riemann problem as suggested by van Leer. The remainder of this section describes the algorithm used to find the exact solution to the Riemann problem. Following this, the exact solution to the particular Riemann problem which is to be solved numerically is shown for the primitive variables.[1]

This solver described blow uses shock relations at the shock and rarefaction relations at a rarefaction. The Riemann solver used by Colella [121] uses shock relations for both types of waves. This results in a much simpler solver. For a detailed look at the Riemann problem see the review paper by Menikoff and Plohr [68].

The algorithm that follows begins from initial data which is defined in two states right, r, and left, l, which are shown graphically in Fig. B.1. The basic algorithm is

[1] The primitive variables are the density, ρ, the velocity, u, the internal energy, ϵ, and the pressure, p.

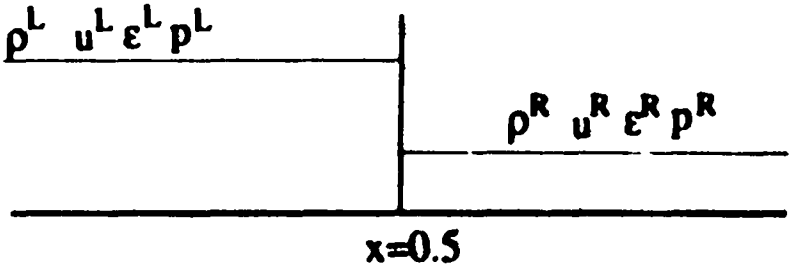


Figure B.1: A representation of the initial conditions for the Riemann Problem.

given below.

Algorithm 6 *[Exact Riemann Solution for a Shock Tube]*

Initial condition, $(\rho_l, u_l, \varepsilon_l, p_l)^T, (\rho_r, u_r, \varepsilon_r, p_r)^T$

$p_* = \frac{1}{2}(\rho_l + \rho_r) + \frac{1}{2}\sqrt{\frac{\gamma(p_l + p_r)}{4}(\rho_l + \rho_r)}\,(u_l + u_r)$

begin

Do While not converged

begin

begin

if $p_* > p_l$ **then**

$$M_l = \sqrt{\gamma\rho_l p_l}\sqrt{\frac{\gamma+1}{2\gamma}\frac{p_*}{p_l} + \frac{\gamma-1}{2\gamma}}$$

$$(u_{l*})' = -\frac{(M_l)^2 + \gamma\rho_l p_l}{2(M_l)^3}$$

else

$$M_l = \frac{\gamma-1}{2\gamma}\sqrt{\gamma\rho_l p_l}\left(1 - \frac{p_*}{p_l}\right)$$

$$(u_{l*})' = -\frac{1}{\gamma\rho_l p_l\left(\frac{p_*}{p_l}\right)^{\frac{\gamma-1}{2\gamma}}}$$

endif

if $p_* > p_r$ **then**

$$M_r = \sqrt{\gamma\rho_r p_r}\sqrt{\frac{\gamma+1}{2\gamma}\frac{p_*}{p_r} + \frac{\gamma-1}{2\gamma}}$$

$$(u_{r*})' = -\frac{(M_l)^2 + \gamma\rho_r p_r}{2(M_r)^3}$$

else

$$M_r = \frac{\gamma-1}{2\gamma}\sqrt{\gamma\rho_l p_l}\left(1 - \frac{p_*}{p_r}\right)$$

$$(u_{r*})' = -\frac{1}{\gamma\rho_r p_r\left(\frac{p_*}{p_r}\right)^{\frac{\gamma-1}{2\gamma}}}$$

endif

$$u_{l*} = u_l - \frac{p_* - p_l}{M_l}$$
$$u_{r*} = u_r - \frac{p_* - p_r}{M_r}$$
$$p_* = p_* - \frac{u_{l*} - u_{r*}}{(u_{l*})' - (u_{r*})'}$$

end

check convergence

end

$$u_* = \frac{1}{2}(u_{l*} + u_{r*})$$

This algorithm was used to produce the solutions shown in Appendix A. These show the characteristics of the exact solution to a Riemann problem for an ideal gas when both sides of the initial condition are at rest and the density and pressure are discontinuous.

B.3.2 Approximate Riemann Solvers

The basis of approximate Riemann solvers is discussed in [40]. For a Riemann solver to be conservative, the following relation should hold assuming Γ is chosen to be large enough

$$\int_{-\Gamma}^{\Gamma} \mathbf{W}(\xi)\, d\xi = \Gamma(\mathbf{U}_l + \mathbf{U}_r) + \mathbf{F}_l - \mathbf{F}_r\,. \tag{B.4a}$$

This relation can be rewritten to give

$$\int_{-\Gamma}^{0} \mathbf{W}(\xi)\, d\xi = \Gamma\mathbf{U}_l + \mathbf{F}_l - \mathbf{F}_{lr}\,, \tag{B.4b}$$

and

$$\int_{0}^{\Gamma} \mathbf{W}(\xi)\, d\xi = \Gamma\mathbf{U}_r + \mathbf{F}_{lr} - \mathbf{F}_r\,. \tag{B.4c}$$

These relations can be manipulated to give various approximate Riemann solvers. For instance choosing $\Gamma = 1/\sigma$ give the Lax-Friedrichs scheme and $\Gamma = \max_k\left(\lambda^k\right)$ gives the Rusanov or Local Lax-Friedrichs scheme.

In this section, six approximate Riemann solutions for the equations in Lagrangian coordinates are given. The solution using these algorithms gives the "solution in the small" in Lagrangian coordinates.

The solution in the small algorithm is identical for all three methods described below except for the details.

Algorithm 7 *[Solution in the Small]*

1. For each grid cell edge, $j + \frac{1}{2}$, compute the right and left variable values from the reconstruction polynomial, $\mathbf{U}_j(x) = \mathbf{P}_j(x)$. Here for Godunov's method $\mathbf{P}_j(x) = \mathbf{U}_j$, thus $\mathbf{U}_l = \mathbf{U}_j$ and $\mathbf{U}_r = \mathbf{U}_{j+1}$.

2. Compute the solution, $\mathbf{U}_{lr}$, (exact or approximate) to the Riemann problem with initial states $\mathbf{U}_l$ and $\mathbf{U}_r$.

3. Use $\mathbf{U}_{lr}$ to compute the flux functions, $\mathbf{F}_{lr}$, at the interface $j + \frac{1}{2}$.

B.3.3 Approximate Riemann Solvers for the Scalar Wave and Burgers' Equation

The scalar wave equation, (A.1a), is the simplest equation to solve. For a constant wave speed, a, the solution to the local Riemann problem is

$$f_{j+\frac{1}{2}} = \frac{a}{2}\left(u_{j+\frac{1}{2},l} + u_{j+\frac{1}{2},r}\right) - \frac{|a|}{2}\left(u_{j+\frac{1}{2},r} - u_{j+\frac{1}{2},l}\right) . \tag{B.5}$$

Here the cell edge values are given by the local interpolating polynomials as

$$u_{j+\frac{1}{2},l} = P_j(x_{j,r}) \tag{B.6a}$$

and

$$u_{j+\frac{1}{2},r} = P_{j+1}(x_{j+1,l}) , \tag{B.6b}$$

where $x_{j,r} = x_{j+1,l} = x_{j+\frac{1}{2}}$.

In [158] van Leer gives the local Riemann solution to Burgers' equation. Given in the nomenclature of this appendix this solution restated is

$$f_{j+\frac{1}{2}} = \max\left[\frac{1}{2}\max\left(u_{j+\frac{1}{2},l},0\right)^2, \frac{1}{2}\min\left(u_{j+\frac{1}{2},r},0\right)^2\right] . \tag{B.7}$$

B.3.4 Naive Riemann Solver

This method of closing the equations is included because it is so frequently done despite a number of deficiencies. The system is considered to be a set of uncoupled equations and the terms are considered one-by-one. If a term appears to advect the variable, such as density in the mass conservation equation or energy in the energy equation, it is upwind differenced. If it is another term, such as the pressure gradient in the momentum equation or the work term ($p\nabla \cdot \mathbf{u}$) in the energy equation, it is centrally differenced.

For the Lagrangian equation set all spatial derivatives are centrally differenced. This is justified because sound waves travel in both directions from an interface, and the effect is nearly correct. This is the manner in which the flux corrected transport

algorithms are implemented [143, 4]. As the magnitude of the jump increases, the errors become larger resulting in unphysical solutions.

B.3.5 Lax-Friedrichs Riemann Solver

The simplest Riemann solver is the Lax-Friedrichs solver [55]. This solver is the most diffusive of the solvers discussed in this appendix. It corresponds to Godunov's method over a staggered grid [200] and has a simple form. The only requirement is that the CFL condition be satisfied. The form of the flux function is

$$\mathbf{F}_{lr} = \frac{1}{2}\left[\mathbf{F}_l + \mathbf{F}_r - \frac{1}{\sigma}(\mathbf{U}_r - \mathbf{U}_l)\right] . \tag{B.8}$$

B.3.6 Local Lax-Friedrichs Riemann Solver

Recently [65, 66, 169], the Lax-Friedrichs scheme has been resurrected. This method is also known more classically as Rusanov's method [233, 189]. It has been changed to a form known as the local Lax-Friedrichs (LLF). In this form, it is less diffusive than the classical form of the Lax-Friedrichs method, but still has the advantage of satisfying the entropy equality. The form of the flux simply depends on the maximum (absolute value) wavespeed locally and the form is given by

$$\mathbf{F}_{lr} = \frac{1}{2}\left[\mathbf{F}_l + \mathbf{F}_r - \eta_{lr}(\mathbf{U}_r - \mathbf{U}_l)\right] , \tag{B.9a}$$

where $\eta_{lr} = \sup_k \left|\lambda^k_{lr}\right|$. For the Lagrangian flow equation, this is equal to

$$\eta_{lr} = \max(C_l, C_r) . \tag{B.9b}$$

B.3.7 HLLE Riemann Solver

In their paper [30], Harten, Lax, and van Leer discuss several approximate Riemann solvers in a theoretical context. One of these solvers is derived for a solution containing the left and right initial state plus one intermediate state. Einfeldt [128] then took this basis and showed how this theoretical construct could be used as a practical approximate Riemann solver. Work on this method has also been done by Davis [189]. This method has several desirable properties: its simplicity, ease of implementation, and satisfaction of entropy inequalities.

The general form of a flux function with this solver is

$$f_{lr} = \frac{b^+_{lr} f(u_l) - b^-_{lr} f(u_r)}{b^+_{lr} - b^-_{lr}} + \frac{b^+_{lr} b^-_{lr}}{b^+_{lr} - b^-_{lr}}(u_r - u_l) , \tag{B.10a}$$

where $b^+_{lr} = \max(0, b^r_{lr})$ and $b^-_{lr} = \min\left(0, b^l_{lr}\right)$. The signal speeds b^r_{lr} and b^l_{lr} are upper and lower bounds on the signal velocities, respectively. Reference [231] makes the

suggestion for the computation of b^r_{lr} and b^l_{lr}. The formulas are

$$b^r_{lr} = \max\left(a_{R,max}, a_{LR,max}\right) , \tag{B.10b}$$

and

$$b^l_{lr} = \min\left(a_{L,min}, a_{LR,min}\right) , \tag{B.10c}$$

where max and min refer to the maximum and minimum characteristic speeds at the respective locations. The values for a_{lr} come from Roe's linearization that is discussed below.

For the Lagrangian flow equations, this leads to a straightforward algorithm. The Lagrangian cell interface fluxes can be written

$$\mathbf{F}_{lr} = \frac{C_{lr}\mathbf{F}(\mathbf{U}_l) + C_{lr}\mathbf{F}(\mathbf{U}_r)}{C_{lr} + C_{lr}} - \frac{C_{lr}C_{lr}}{C_{lr} + C_{lr}}(\mathbf{U}_r - \mathbf{U}_l) , \tag{B.10d}$$

which simplifies to

$$\mathbf{F}_{lr} = \frac{1}{2}\left[(\mathbf{F}(\mathbf{U}_l) + \mathbf{F}(\mathbf{U}_r)) - C_{lr}(\mathbf{U}_r - \mathbf{U}_l)\right] , \tag{B.10e}$$

with b^+_{lr} in eq. (B.10a) being replaced by C_{lr}, the largest signal speed, and b^-_{lr} being replaced by $-C_{lr}$, the smallest signal speed.

B.3.8 Roe's Riemann Solver

Roe presented this solver in [232] and the derivation given below gives the same results. The main difference is that the form given here is useful in the derivation of the flux splitting scheme. Roe's approximate Riemann solver uses the Jacobian of the flux function to derive a characteristic decomposition of the system of equations; thus in general

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0 \Rightarrow \frac{\partial \mathbf{U}}{\partial t} + A\frac{\partial \mathbf{U}}{\partial x} = 0 , \tag{B.11a}$$

where $A = \partial\mathbf{F}/\partial\mathbf{U}$ is the Jacobian matrix. If I define the decomposition as

$$A = R\Lambda R^{-1} ,$$

Λ is a diagonal matrix with the eigenvalues of A on the diagonal, R is the matrix of right eigenvectors (columns), and R^{-1} is the matrix left eigenvectors (rows). Characteristic equations are then defined as

$$\frac{\partial \alpha}{\partial t} + \Lambda\frac{\partial \alpha}{\partial x} = 0 , \tag{B.11b}$$

where $\alpha = R^{-1}U$. These equations can be solved with upwind biased methods to get physically correct propagation of information for data associated with each separate wave.

For a scalar wave equation, the expression for an upwind biased flux can be written as,

$$f_{l_r} = \frac{1}{2}\left[(f_l + f_r) - |a_{l_r}|\,(u_r - u_l)\right] , \tag{B.11c}$$

where L and R refer to the states to the left and right of the cell interface $j + \frac{1}{2}$. For Roe's Riemann solver and a system of equations, the flux can be expressed as

$$\mathbf{F}_{l_r} = \frac{1}{2}\left[(\mathbf{F}_l + \mathbf{F}_r) - \sum_k \mathbf{r}^k_{l_r}\left|\mathbf{a}^k_{l_r}\right|(\alpha_r - \alpha_l)\right] , \tag{B.11d}$$

where r^k is the k^{th} right eigenvector and

$$\alpha_r = \mathbf{l}^k_{l_r} \cdot \mathbf{U}_r .$$

Roe defined the Jacobian to be used in this numerical approximation to have the property

$$\mathbf{F}_r - \mathbf{F}_l = A\,(\mathbf{U}_r - \mathbf{U}_l) \tag{B.11e}$$

for averaging the values to find A. For the Euler equations, the averaging procedure is somewhat more complicated than simple averaging, but for the fluid equations in Lagrangian coordinates simple averaging suffices. Therefore, the following relations are used:

$$p_{l_r} = \frac{1}{2}(p_l + p_r) , \tag{B.12a}$$

$$\tau_{l_r} = \frac{1}{2}(\tau_l + \tau_r) , \tag{B.12b}$$

and

$$C^2_{l_r} = \frac{\gamma p_{l_r}}{\tau_{l_r}} . \tag{B.12c}$$

When λ^k can change sign, one slight modification of the above methodology is used for nonlinear equations and systems; as suggested by Yee [134] an entropy fix is implemented for the donor-cell differencing, which modifies the use of the absolute value in donor-cell differencing of a characteristic speed, by

$$\psi(z) = \begin{cases} |z| & \text{if } |z| \geq \epsilon \\ (z^2 + \epsilon^2)/2\epsilon & \text{if } |z| < \epsilon \end{cases} , \tag{B.13}$$

if one is dealing with a linear equation set $\epsilon = 0$. The parameter ϵ is determined by

the following equation [30],

$$\epsilon = \max\left[0, a_{j+\frac{1}{2}} - a_j, a_{j+1} - a_{j+\frac{1}{2}}\right] .$$

This averaging for the Euler equations requires that a parameter be defined by

$$D_{j+\frac{1}{2}} = (\rho_{j+1}/\rho_j)^{1/2} , \tag{B.14a}$$

which is in turn used to define the following cell edge values:

$$u_{j+\frac{1}{2}} = \frac{D_{j+\frac{1}{2}} u_{j+1} + u_j}{D_{j+\frac{1}{2}} + 1} , \tag{B.14b}$$

$$H_{j+\frac{1}{2}} = \frac{D_{j+\frac{1}{2}} H_{j+1} + H_j}{D_{j+\frac{1}{2}} + 1} . \tag{B.14c}$$

and

$$c_{j+\frac{1}{2}} = \left[(\gamma - 1)\left(H_{j+\frac{1}{2}} - \frac{1}{2}u^2_{j+\frac{1}{2}}\right)\right]^{1/2} , \tag{B.14d}$$

where

$$H = \frac{\gamma p}{(\gamma - 1)\rho} + \frac{1}{2}u^2 . \tag{B.14e}$$

For the Euler equations, the eigenvalues of the flux Jacobian are

$$\left(a^1, a^2, a^3\right) = (u - c, u, u + c) . \tag{B.15a}$$

The right eigenvectors form a matrix

$$R = \left(r^1, r^2, r^3\right) = \begin{bmatrix} 1 & 1 & 1 \\ u - c & u & u + c \\ H - uc & \frac{1}{2}u^2 & H + uc \end{bmatrix} , \tag{B.15b}$$

and by using

$$z_1 = \frac{1}{2}(\gamma - 1)\frac{u^2}{c^2} ,$$

$$z_2 = \frac{\gamma - 1}{c^2} .$$

the left eigenvectors form a matrix

$$R^{-1} = \begin{bmatrix} l^1 \\ l^2 \\ l^3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\left(z_1 + \frac{u}{c}\right) & -\frac{1}{2}\left(z_2 u + \frac{1}{c}\right) & \frac{1}{2}z_2 \\ 1 - z_1 & z_2 u & -z_2 \\ \frac{1}{2}\left(z_1 - \frac{u}{c}\right) & -\frac{1}{2}\left(z_2 u - \frac{1}{c}\right) & \frac{1}{2}z_2 \end{bmatrix} . \qquad \text{(B.15c)}$$

For the Lagrangian flow eqns. (B.2a)–(B.2c), the flux Jacobian is

$$A = \begin{bmatrix} 0 & -1 & 0 \\ -C^2/\gamma & -u(\gamma-1)/\tau & (\gamma-1)/\tau \\ -uC^2/\gamma & \tau C^2/\gamma - u^2(\gamma-1)/\tau & u(\gamma-1)/\tau \end{bmatrix} , \qquad \text{(B.16a)}$$

using an ideal gas equation of state. As stated before the matrix has the eigenvalues of $-C$, 0, and C, and the corresponding right eigenvectors are

$$R = \begin{bmatrix} 1 & 1 & 1 \\ C & 0 & -C \\ uC - p & p/(\gamma-1) & -uC - p \end{bmatrix} ; \qquad \text{(B.16b)}$$

the left eigenvectors are

$$R^{-1} = \begin{bmatrix} \frac{1}{2\gamma} & \frac{1}{2C} + \frac{u(\gamma-1)}{2\gamma p} & \frac{1-\gamma}{2\gamma p} \\ \frac{\gamma-1}{\gamma} & \frac{u(1-\gamma)}{\gamma p} & \frac{\gamma-1}{\gamma p} \\ \frac{1}{2\gamma} & -\frac{1}{2C} + \frac{u(\gamma-1)}{2\gamma p} & \frac{1-\gamma}{2\gamma p} \end{bmatrix} . \qquad \text{(B.16c)}$$

These matrices and the definition of the flux functions above eq. (B.11d) give the method for solution. Roe [232] noted that the actual implementation for this case is somewhat simpler because of the great amount of cancelation of terms as they are multiplied out. The flux vector gives the simplification

$$\mathrm{F} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} u_* \\ p_* \\ p_* u_* \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(u_l + u_r) - \frac{1}{2}(p_r - p_l)/C_{lr} \\ \frac{1}{2}(p_l + p_r) - \frac{1}{2}(u_r - u_l) C_{lr} \\ p_* u_* \end{bmatrix} . \qquad \text{(B.16d)}$$

An interesting footnote to this discussion is that these expressions were developed by Richtmyer and Morton [31, pages 342-345] as a linearized version of Godunov's

method. This method was related to the work of Courant, Issacson, and Rees [54] in order to draw a direct analogy between that method and Godunov's original work.

As is shown below, the right eigenvectors are useful in the derivation flux splittings for these equations. The full matrices also may be useful in visualizing the extension of this method to real gases with more general equations of state.

B.3.9 The Engquist-Osher Solver

The Engquist-Osher Riemann solver [127] has a number of useful properties. It is somewhat different than the others presented here. The scheme takes into account the effects of sonic points and thus satisfies entropy constraints. It is built upon the knowledge that there are a finite number of jumps to states, which can be determined by the characteristic decomposition (Riemann invariants) of the problem. Given these jumps, a well-defined path of integration can be defined for a system.

The form for the fluxes [104] is

$$f_{lr}^{EO} = f(u_*) - \int_{u_l}^{u_*} \max(a(s),0)\ ds + \int_{u_*}^{u_r} \min a(s),0\ dx\ , \tag{B.17}$$

where u_* is a reference state and $a(s)$ is the characteristic speed as a function of position in phase space. It is generally wise to choose u_* to be one of the states at the cell edge. Using the definitions of the characteristic decomposition used for Roe's solver, the fluxes for a system can be written

$$\mathbf{F}_{lr}^{EO} = \mathbf{F}(\mathbf{U}_l) + \sum_{k=1}^{N} \left(\int_0^{\alpha^k} \min\left(\lambda^k, 0\right)\ d\alpha \right) \cdot \mathbf{r}^k\ . \tag{B.18}$$

In this form, the functions to the right of $\mathbf{F}(\mathbf{U}_l)$ only have to be evaluated if the sign of λ_k becomes negative (indicating a change of direction in the upwinding). This change can happen during any of the jumps defined by the Riemann invariants. Because the eigenvalues of the Lagrangian flow equations do not contain sonic points, the effects of this solver are not profound when compared to Roe's Riemann solver). For all intents and purposes, the results obtained with this solver are nearly identical to those obtained with Roe's solver. The integration procedure adds some additional numerical dissipation to the solver not found with Roe's solver.

B.3.10 Flux Splitting

The third approximate Riemann solver used is the process of flux splitting [125]. This method has been widely used for the Euler equations. For the Euler equations, the process of flux splitting has some difficulties because the characteristics can change sign at sonic points (which motivated van Leer's work [126] on flux splitting methods), thus forcing the basis of the algorithm to take this behavior into account. This also

can create difficulty in meeting entropy requirements (also a problem for Roe's solver for the Euler equations). Part of the beauty of the Lagrangian equations is that each characteristic does not change sign; thus the scheme is the same for every grid point, thereby reflecting the symmetry of the system that led to all the cancellation in the final result of the section describing Roe's method.

Upwind differencing in its most basic form is the basis of flux splitting. In general, upwind differencing can be defined for a scalar advection law as follows:

$$u_j^{n+1} = u_j^n - \lambda\left(f^+_{j+\frac{1}{2}} - f^+_{j-\frac{1}{2}}\right) - \lambda\left(f^-_{j+\frac{1}{2}} - f^-_{j-\frac{1}{2}}\right) . \tag{B.19a}$$

where

$$f^+ = \max(f_l, 0) \ , \text{ and } f^- = \min(f_r, 0) \ . \tag{B.19b}$$

This general concept can be extended to systems of equations by the type of decomposition described in the previous section. Given the eigenvalues of the system that define the direction of the flow for a set of characteristic variables, a flux splitting can be defined. For this purpose, the eigenvalues are split as

$$\Lambda = \Lambda^- + \Lambda^+ \ , \tag{B.19c}$$

and the flux Jacobian is split accordingly as

$$A = A^- + A^+ \ , \tag{B.19d}$$

with each matrix corresponding to the appropriate eigenvalue direction. These matrices can be constructed under the condition that

$$\mathbf{F} = A\mathbf{U} \ ; \tag{B.19e}$$

thus,

$$\mathbf{F}^-_{lr} = A^-\mathbf{U}_r \ , \text{ and } \mathbf{F}^+_{lr} = A^+\mathbf{U}_l \ . \tag{B.19f}$$

To derive the flux splitting used here, I draw on an observation reported in [35] that the flux splittings can be found through the right eigenvectors of the flux Jacobian. Using the results of the previous section, the following equation set can be constructed

$$\lambda_1\beta_1\begin{bmatrix} 1 \\ C \\ uC - p \end{bmatrix} + \lambda_2\beta_2\begin{bmatrix} 1 \\ 0 \\ p/(\gamma - 1) \end{bmatrix} + \lambda_3\beta_3\begin{bmatrix} 1 \\ -C \\ -uC - p \end{bmatrix} = \begin{bmatrix} -u \\ p \\ up \end{bmatrix} \ , \tag{B.20a}$$

where

$$\lambda = \begin{bmatrix} -C \\ 0 \\ C \end{bmatrix}.$$

This equation set can be solved to yield the appropriate flux splitting.

The resulting flux splitting is

$$\mathbf{F}_{lr}^{-} = \begin{bmatrix} \frac{1}{2}\left(\frac{p_r}{C_{lr}} - u_r\right) \\ \frac{1}{2}\left(p_r - u_r C_{lr}\right) \\ \frac{1}{4}\left(\frac{p_r}{C_{lr}} - u_r\right)\left(u_r C_{lr} - p_r\right) \end{bmatrix} \tag{B.20b}$$

and

$$\mathbf{F}_{lr}^{+} = \begin{bmatrix} -\frac{1}{2}\left(\frac{p_l}{C_{lr}} + u_l\right) \\ \frac{1}{2}\left(p_l + u_l C_{lr}\right) \\ \frac{1}{4}\left(\frac{p_l}{C_{lr}} + u_l\right)\left(u_l C_{lr} + p_l\right) \end{bmatrix}, \tag{B.20c}$$

where C_{lr} is the Roe averaged sound speed. Close inspection of the above expressions reveals that the energy flux in each case is similar to Roe's solver in that $F_3 = -p_* u_*$. Still closer inspection reveals that this flux splitting is in fact identical to Roe's solver.

Remark 26 *The use of the HLLE or LLF solvers promises to significantly ease the implementation of Godunov type schemes with Riemann solvers. This is especially true for complex systems of equations or for implicit algorithms. If maximum and/or minimum wavespeeds cannot be found, then by using an estimate plus (or minus) some constant, which is large enough (this constant or estimate must be used in computing stability limits), a physical solution can be found. The one problem of this approach is that the solutions found with these approaches can be significantly more diffused than Roe's algorithm (as shown in the following section).*

B.4 Results

In this section, the results obtained through the use of the algorithms described above is given and discussed. Several test problems taken from the literature are used: Sod's problem [41], Lax's problem [55], and a blast wave problem [44]. In each case, only the solution for density is given for brevity. This should not present too much of a detriment because the density profile in each problem captures the essence of each method's strengths and weaknesses. For Lax's and Sod's problem, an exact solution is used to provide and absolute comparison of the results. For the blast wave problem,

no exact solution exists; thus, for an absolute comparison, a converged high-resolution second-order solution is used.[2]

For brevity, the examination of the solution's properties is done using the density profile obtained. The density is an effective measure of algorithmic performance because it contains all the pertinent structures in the one-dimensional flow (shocks, rarefactions, and contact discontinuities).

B.4.1 Sod's Problem

Figure B.2a shows the solution obtained through the use of the naive Riemann solver with Godunov's method. The most noticeable feature of this plot is the oscillations behind the shock ($X \approx 85$). The shock is relatively sharp, but the contact discontinuity is smeared severely. Less notable is the small oscillation ahead of the rarefaction wave as well as what appears to be a small expansion shock in the rarefaction wave ($X \approx 30$). These oscillations can be reduced significantly by reducing the time step used in the calculations (which increases the inherent dissipation in the solution). In general, the solution by this method is unsatisfactory.

In Figs. B.2b and B.2c the solutions found with Roe's and Engquist-Osher's Riemann solver are given. These solutions are nearly identical with Engquist-Osher's Riemann solver, but have slightly more smearing. The shock is about four cells wide in both cases, but is slightly sharper with Roe's method. The contact discontinuity and the rarefaction wave are both smeared significantly, but the solution appears to be physical throughout the domain for both methods. It should be noted that Engquist-Osher's Riemann solver is more expensive than Roe's Riemann solver

Figure B.2d shows the results for the HLLE Riemann solver and Fig. B.2e shows those for the LLF Riemann solver. Both of these solutions show a great deal more numerical diffusion in the rarefaction wave through the contact discontinuity. The HLLE Riemann solver gives a crisp shock wave across approximately two cells. The LLF Riemann solver shows about the same resolution of the shock as Roe's and Engquist-Osher's Riemann solvers. Another notable feature of these solvers is their cost. Both are somewhat cheaper than the more complex Riemann solvers like Roe's and the Engquist-Osher. As with the previous two solvers, the solution is physical in nature.

B.4.2 Lax's Problem

The naive Riemann solver again produces less than satisfactory results. The oscillations behind the shock are present again, but large oscillations are also present between the rarefaction and contact discontinuity. Again there is some semblance of

[2]This is a second-order Godunov method using Roe's superbee [176] flux limiter to enhance the resolution of contact discontinuities and high-resolution limiters for the genuinely nonlinear fields.

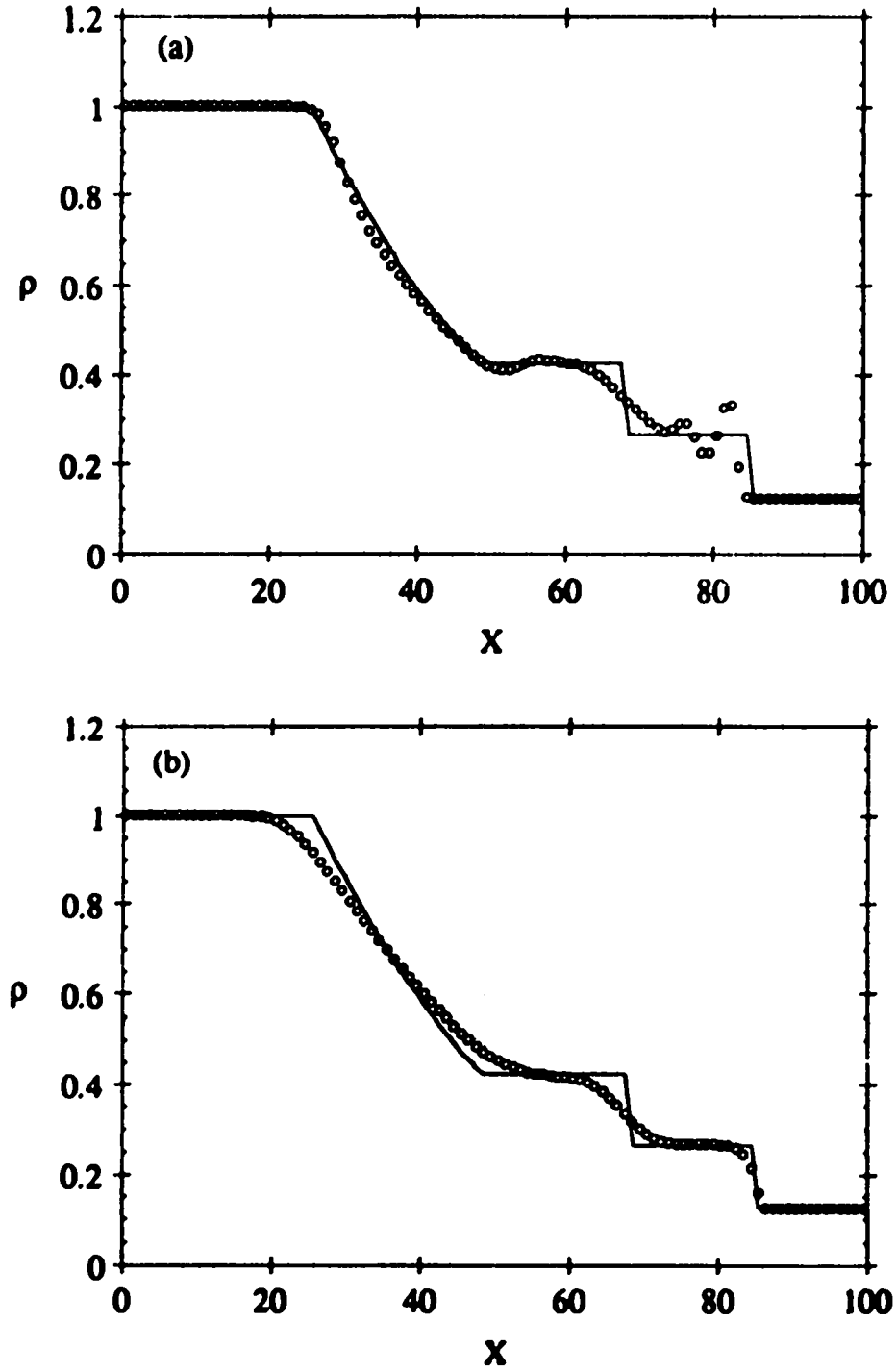
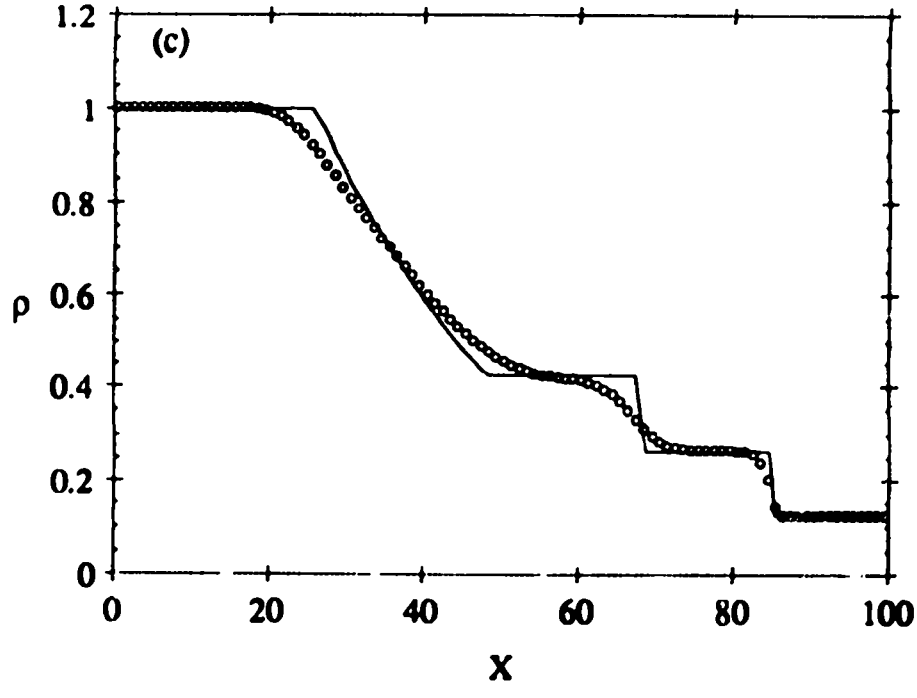


Figure B.2: The solution for Sod's shock tube problem at $t = 20$ is obtained with each of the methods discussed in this appendix. The exact solution is denoted by the solid line in each plot, and the solution obtained with Godunov's method is shown by the circles. Figure B.2a shows the solution obtained with the naive Riemann solver followed by Roe's Riemann solver (B.2b), Engquist-Osher's Riemann solver (B.2c), the HLLE Riemann solver (B.2d) and the LLF Riemann solver (B.2e).



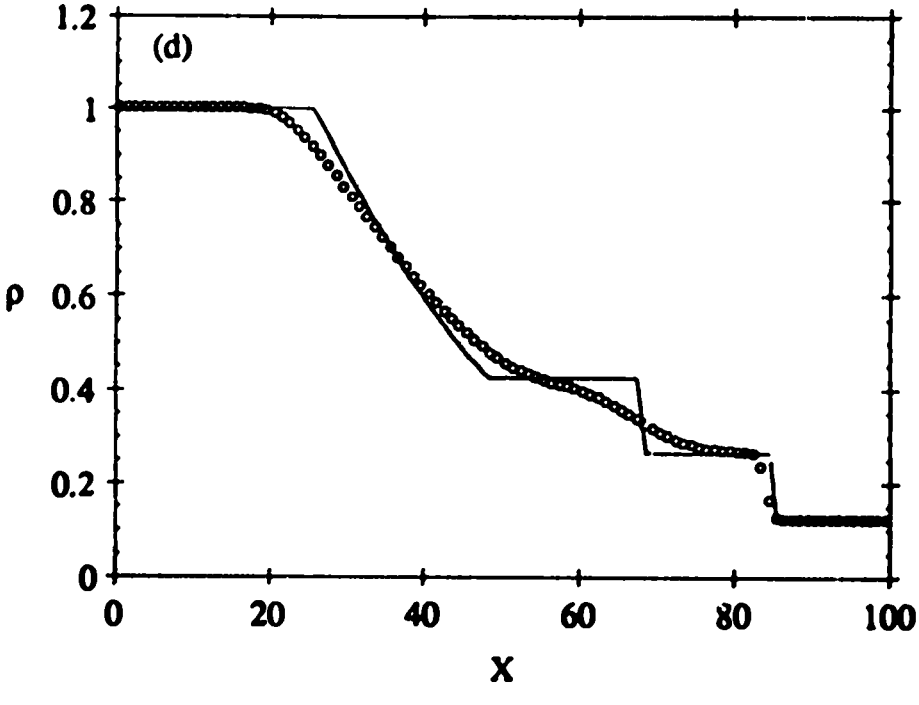


Figure B.2: continued

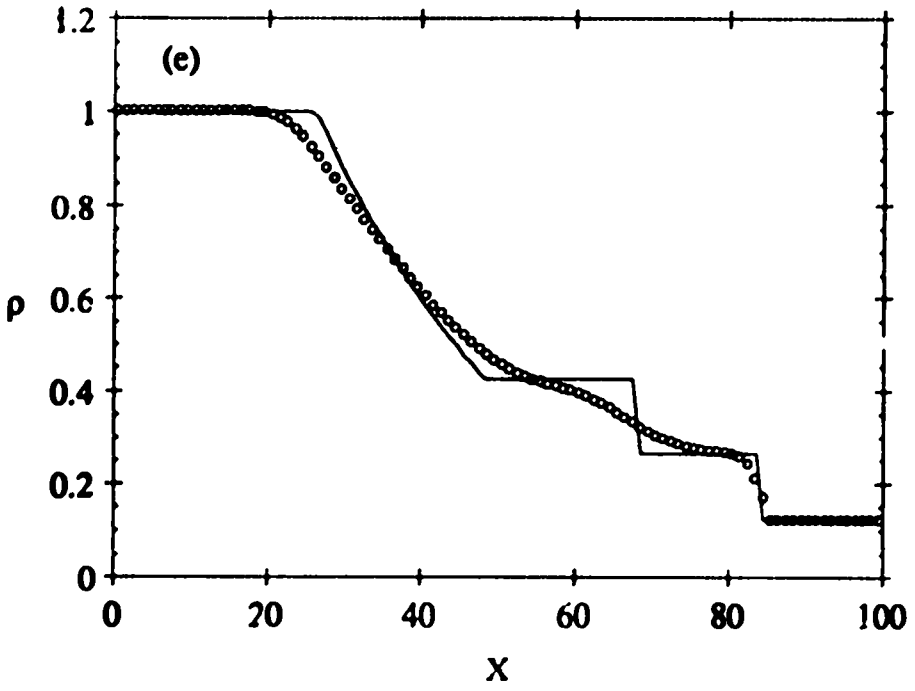


Figure B.2: continued

an expansion shock in the rarefaction wave. Figure B.3a shows these results. The negative features in the solution are gradually removed from the flow as the CFL number is reduced.

The Roe and the Engquist-Osher Riemann solvers again produce nearly identical solutions with the only difference being the slight increase in numerical dissipation for Engquist-Osher's Riemann solver. The shock is smeared to be quite wide as is the contact discontinuity. Figures B.3b and B.3c show that in both cases the rarefaction is smeared. In addition, both solutions slightly clip the square peak in the density profile.

Figures B.3d and B.3e show the HLLE and LLF Riemann solvers respectively. As before, the shock is crisper with the HLLE Riemann solver than either the Roe or Engquist-Osher Riemann solvers, but the clipping of the density peak is more pronounced and the smearing in both the rarefaction wave and contact discontinuity is more severe. The LLF Riemann solver shows the same characteristics, but does not have a crisper shock wave, and the smearing is more severe than that found with the HLLE Riemann solver.

B.4.3 Blast Wave Problem

Figure B.4a shows the results using the naive Riemann solver. The smooth portion of the flow on the left is severely polluted with instabilities as is the shock wave at $X \approx 64$. Other smaller oscillations can be seen past the shock at $X \approx 85$ and next

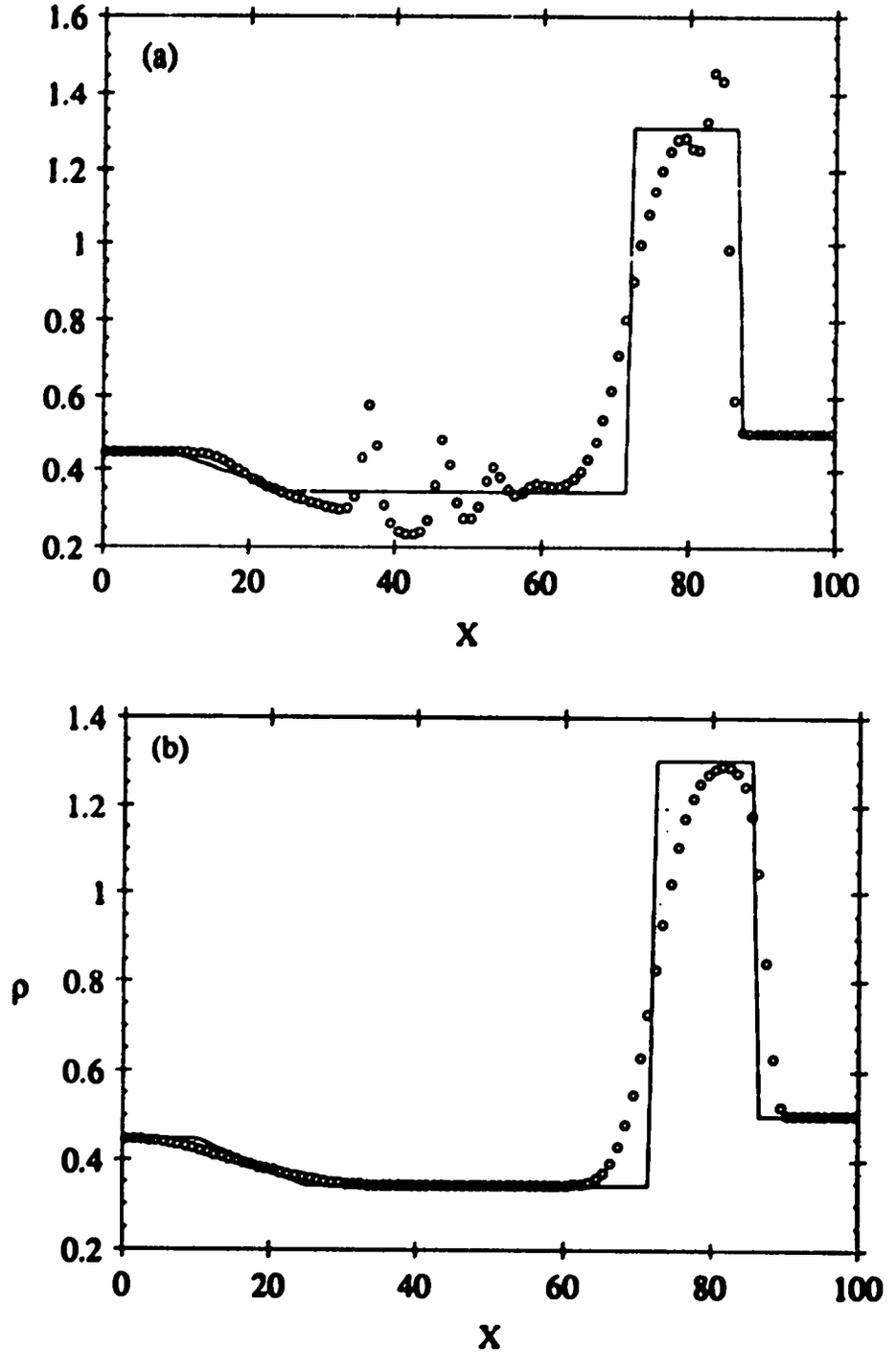


Figure B.3: The solution for Lax's shock tube problem at $t = 15$ is obtained with each of the methods discussed in this appendix. The exact solution is denoted by the solid line in each plot, and the solution obtained with Godunov's method is shown by the circles. Figure B.3a shows the solution obtained with the naive Riemann solve followed by Roe's Riemann solver (B.3b), Engquist-Osher's Riemann solver (B.3c), the HLLE Riemann solver (B.3d), and the LLF Riemann solver(B.3e).

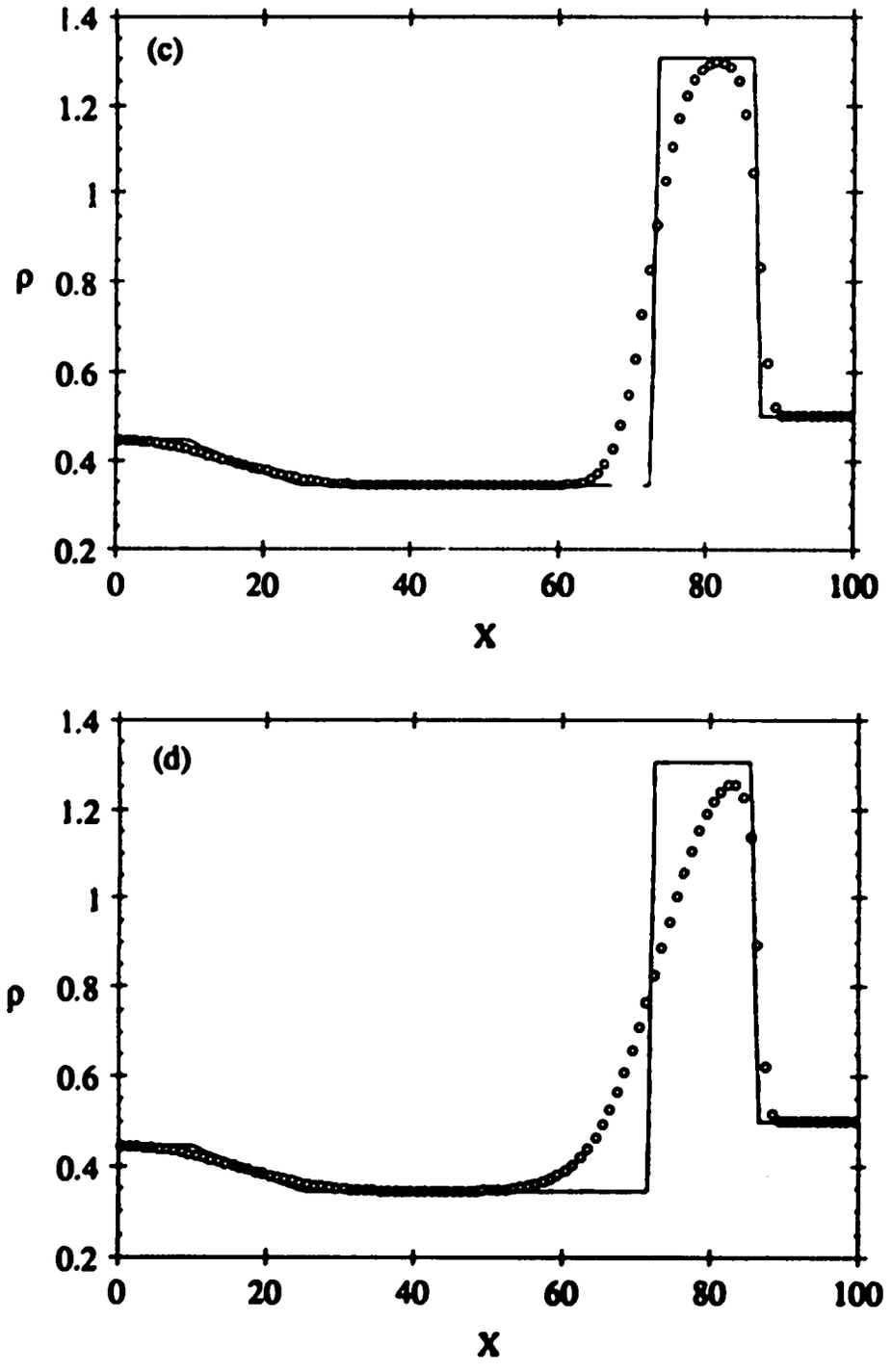


Figure B.3: continued

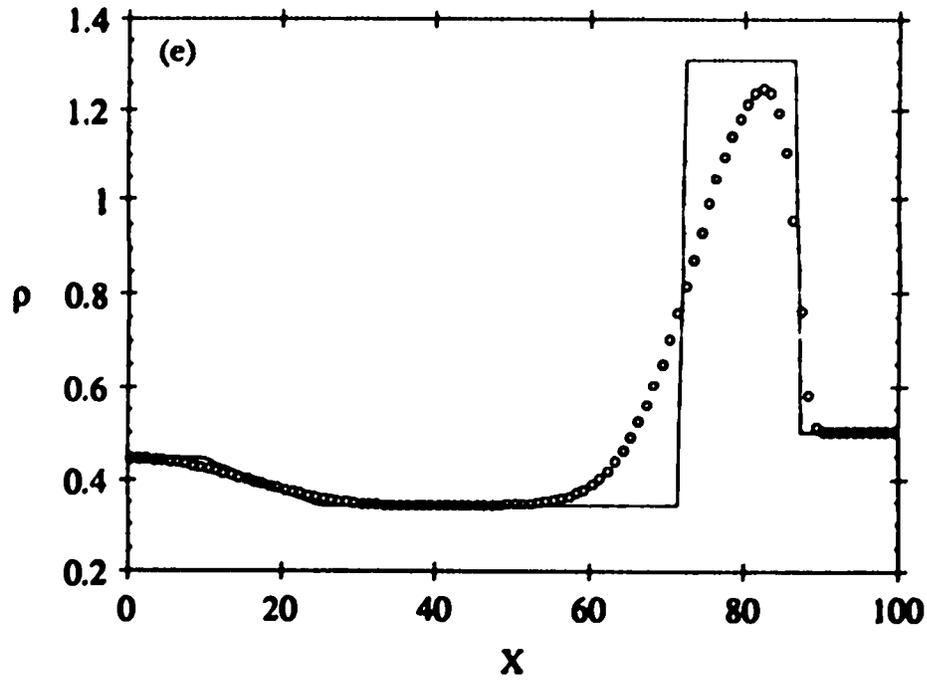


Figure B.3: continued

to the right wall. Although the solution captures some of the essence of the flow, the characteristics of this solution do not indicate that this procedure is robust. Reducing the CFL as before improves the results; however, the improvement is not as quick as with the simpler shock tube type problems.

Figures B.4b and B.4c show the results obtained with the Roe and Engquist-Osher Riemann solvers. As before, these are nearly identical, but Engquist-Osher's Riemann solver degrades the solution peaks slightly more than Roe's. In general, all features of the solution are smeared considerably by the solution procedure. The contact discontinuities at $X \approx 60$ and $X \approx 80$ are both smeared considerably with the first one being totally obscured. The "dip" between the peaks associated with a rarefaction wave is filled in to a large degree.

The results obtained with the HLLE and LLF Riemann solvers are even more diffusive as one might expect. The peaks are clipped to a larger degree and the "dip" between them is filled in to a greater degree. Again the LLF Riemann solver exhibits more dissipation than the HLLE solver, although their performance is nearly indistinguishable. The HLLE Riemann solver also produces a slightly sharper shock at $X \approx 88$ than the other methods (except the naive Riemann solver), although this result is barely perceptible from the figures (Figs. B.4d and B.4e).

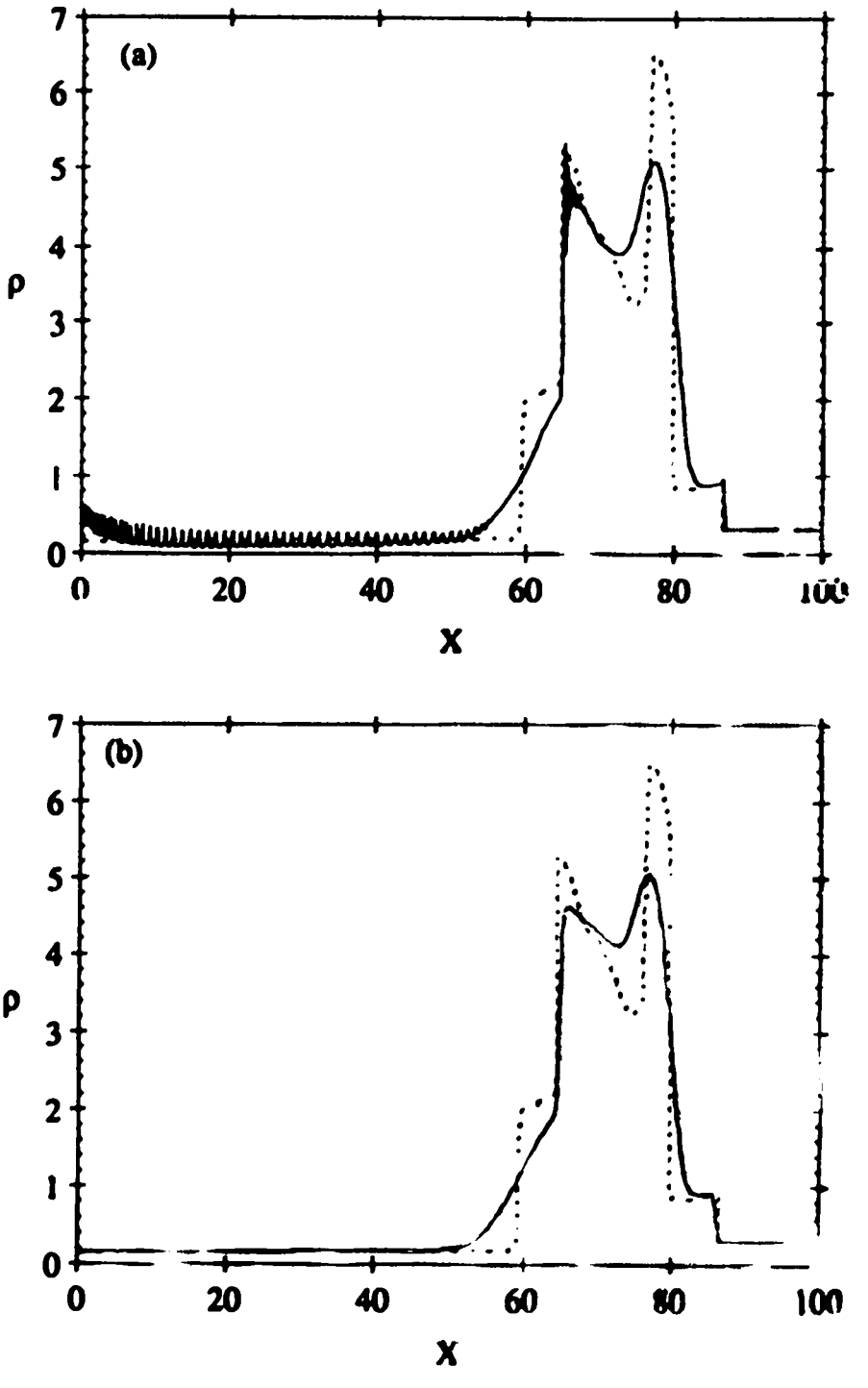


Figure B.4: The solutions to the blast wave problem at $t = 3.8$ are shown. The converged numerical solution is shown by the dashed line and the solid line shows the solution obtained with the approximate Riemann solvers in conjunction with a first-order Godunov method. Figure B.4a shows the solution obtained with the naive Riemann solve followed by Roe's Riemann solver (B.4b), the Engquist-Osher's Riemann solver (B.4c), the HLLE Riemann solver (B.4d), and the LLF Riemann solver(B.4e).

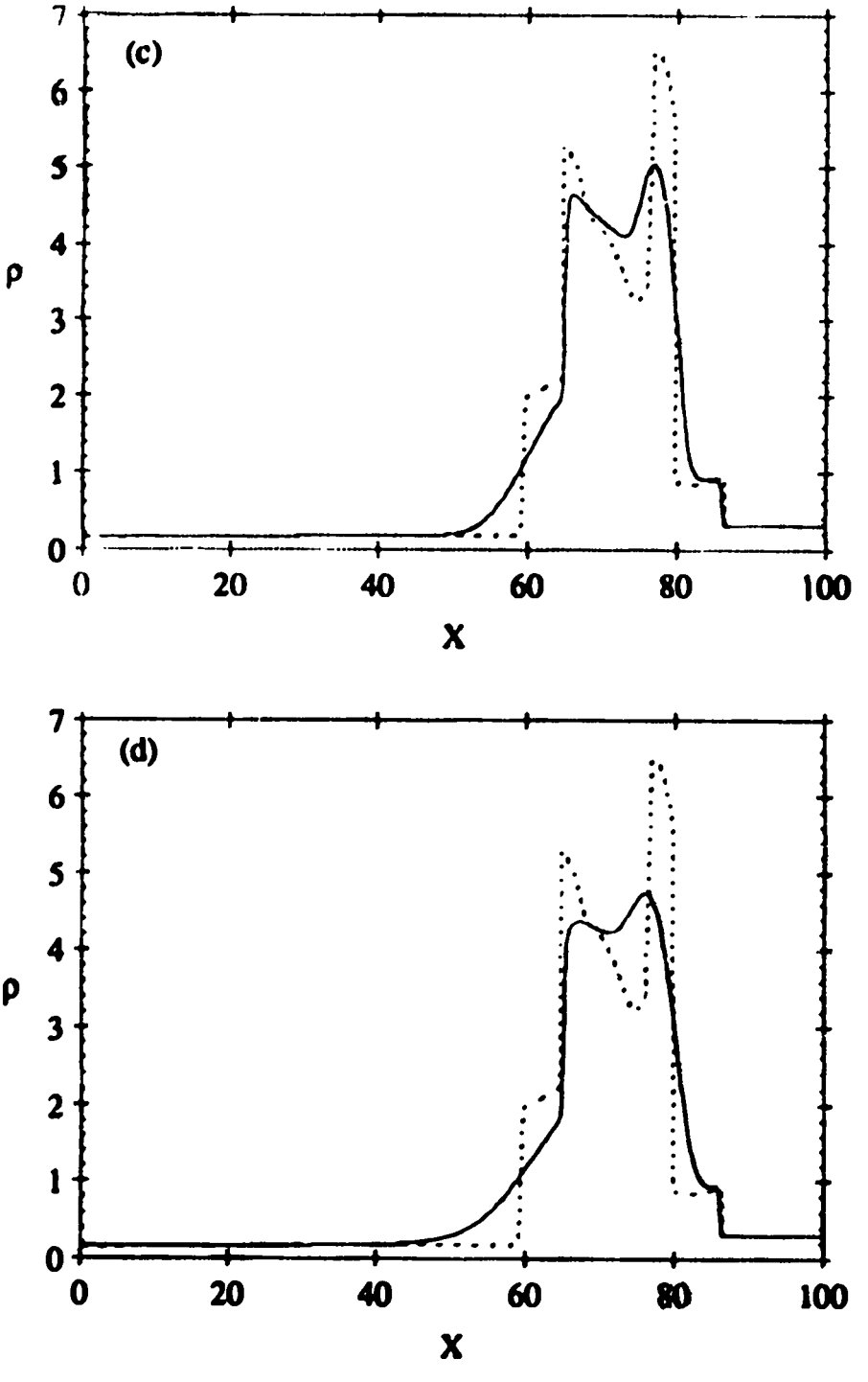


Figure B.4: continued

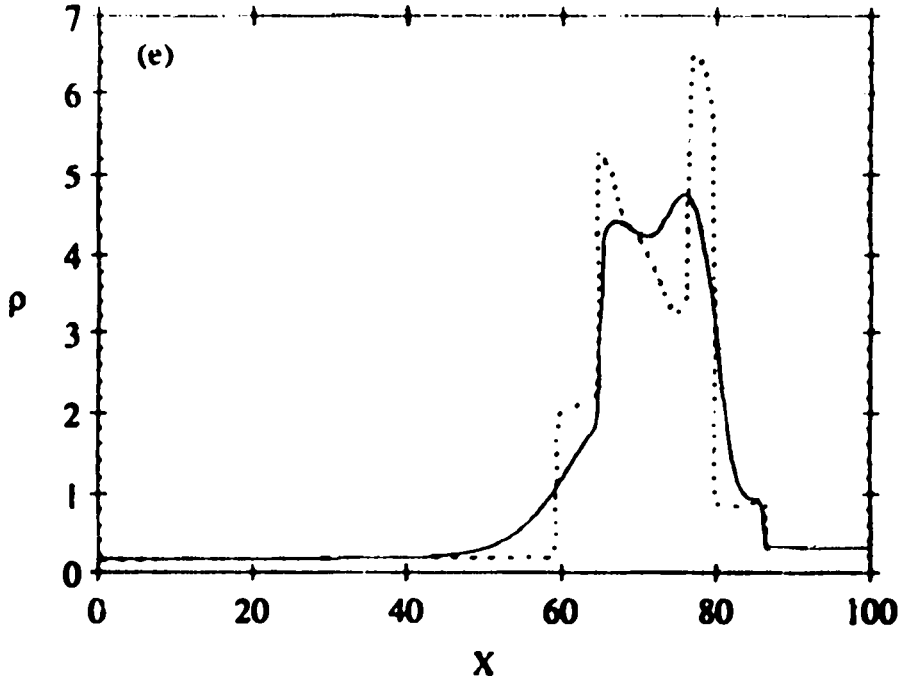


Figure B.4: continued

B.5 Concluding Remarks

This appendix has given the form of various approximate Riemann solvers that may be useful in producing quality results with Godunov's method in Lagrangian coordinates with or without an Eulerian remap. In addition, the results show some of the problems with taking the naive Riemann solver approach. The three test problems show that the other types of Riemann solvers produce physical results (and importantly at a lower cost than "exact" Riemann solvers).

The Roe and Engquist-Osher Riemann solvers both employ a great deal of knowledge of the wave structure of the equation set and as such produce relatively good results. If the wave structure is not as well defined or known, the HLLE and LLF Riemann solvers provide a simple alternative provided good estimates of the wavespeeds present are available. The latter two solvers also are less computationally intensive and generally simpler, and thus offer some saving in that regard.

With the use of higher order "monotone" interpolation principles with the methods given in this appendix, the results for all methods improve.

The extension of high-order methods to systems of equations is explored in the following appendix.

Appendix C.

Extension of High Resolution Schemes to Systems of Conservation Laws

C.1 Introduction

In recent years, there has been an abundance of work deriving high-resolution schemes for hyperbolic conservation laws. Most of the development is made with scalar equations and generalized in some fashion to nonlinear equations or systems of equations. Typically, the extension to systems of equations takes on great importance as is the case with the solution of the Euler equations of compressible flow. Much of the development of high-resolution methods is devoted to the solution of systems of equations as their primary practical use.

This appendix is divided into five sections. The following section introduces the methods used for a scalar wave equation. In the third section, each of these methods is extended to systems of equations. The fourth section presents and discusses results found using these methods for the Euler equations. Finally, concluding remarks are found in the last section. An appendix describes the characteristic decomposition for both conserved and primitive variables.

C.2 Preliminaries

In this appendix, I concentrate my efforts on one specific method and its extension to systems of equations. This method is a standard second-order HOG method augmented with TVD limiters (Chapter 8 and [132]). As noted in [64, 147], the process of solving a problem with a Godunov-type method can be divided into two basic steps: reconstruction or projection and evolution. The evolution step involves the use of some sort of exact or approximate Riemann solvers (see for example Appendix B or [30]). The issue at hand here is the method of projection for systems of equations.

The projection step requires that a piecewise polynomial (or some functional representation) be defined for each cell of the system to reconstruct the variables distribution in space to some level of desired accuracy. In this appendix, the following form is used for this polynomial

$$P_j(x) = n_j + \bar{\Delta}_j n \frac{(x - x_j)}{\Delta_j x} \quad , \quad x \in \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right] \quad , \tag{C.1a}$$

where

$$\widetilde{\Delta_j} u = Q(1,r)\,\Delta_{j-\frac{1}{2}} u\ , \tag{C.1b}$$

with

$$\Delta_{j-\frac{1}{2}} u = u_j - u_{j-1}\ . \tag{C.1c}$$

The mesh spacing is $\Delta x = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$, $x_j = \left(x_{j+\frac{1}{2}} + x_{j-\frac{1}{2}}\right)/2$ and $r = \Delta_{j+\frac{1}{2}} u / \Delta_{j-\frac{1}{2}} u$. The function $Q(1,r)$ is a limiter.

The limiters used in this appendix are discussed in Chapter 8.

The polynomial is then used to define left and right states of the variables at each cell edge, u_l and u_r. These quantities are then used to determine a cell-edge numerical flux f_{lr} via a Riemann solver. In the cases (except one as explained in Section C.4.3) considered in this appendix, Roe's approximate Riemann solver [63] is used. This gives an overall conservative numerical scheme of

$$u_j^{n+1} = u_j^n - \sigma\left(f_{j+\frac{1}{2},lr} - f_{j-\frac{1}{2},lr}\right)\ , \tag{C.2a}$$

with

$$f_{j+\frac{1}{2},lr} = \frac{1}{\Delta t}\int_t^{t+\Delta t} f\left(u\left(x_{j+\frac{1}{2}}, \tau\right)\right) d\tau\ . \tag{C.2b}$$

For extension to systems not using a characteristic decomposition it is likely that other approximate Riemann solvers will be used.

C.2.1 Lax-Wendroff-Type Differencing

Another issue easily addressed with simple model problems is time accuracy. For a second-order accurate scheme spatially, it is often important to attain second-order accuracy temporally. A common practice is to use a Lax-Wendroff approach to time accuracy. From one point of view this reduces to characteristic tracing at the cell edges to get a time-centered estimate of the cell-edge state. For this numerical scheme this yields the following form for cell edge states:

$$u_{j+\frac{1}{2},l}^{n+\frac{1}{2}} = u_j + \frac{1}{2}\widetilde{\Delta_j} u\,(1 - \eta_{lr})\ , \tag{C.3a}$$

and

$$u_{j+\frac{1}{2},r}^{n+\frac{1}{2}} = u_{j+1} - \frac{1}{2}\widetilde{\Delta_{j+1}} u_{j+1}\,(1 + \eta_{lr})\ , \tag{C.3b}$$

$= \sigma$ This can also be viewed as evaluating in the integral in (C.2b) by a

... comparison is shown in Fig. 4.8.

C.2.2 Two-Step Formulation

This procedure becomes more difficult when systems of equations are considered. To combat this difficulty, a procedure in the spirit of the two-step Lax-Wendroff

scheme [114, 113], has been used [159, 158]. The left and right states are computed from the projective polynomial and then used to produce time-centered estimates for the cell-edge states. Given the cell-edge states, $u^n_{j+\frac{1}{2},l}$ and $u^n_{j+\frac{1}{2},r}$, computed with a high-order method, the time-centered estimates are

$$u^{n+\frac{1}{2}}_{j+\frac{1}{2},l} = u^n_{j+\frac{1}{2},l} - \frac{\sigma}{2}\left[f\left(u^n_{j+\frac{1}{2},l}\right) - f\left(u^n_{j-\frac{1}{2},r}\right)\right] , \tag{C.4a}$$

and

$$u^{n+\frac{1}{2}}_{j+\frac{1}{2},r} = u^n_{j+\frac{1}{2},r} - \frac{\sigma}{2}\left[f\left(u^n_{j+\frac{1}{2},l}\right) - f\left(u^n_{j+\frac{1}{2},r}\right)\right] . \tag{C.4b}$$

This gives second-order temporal accuracy and is equivalent to the Lax-Wendroff type procedure for scalar equations.

Remark 27 *Davis [189] presents an alternate two-step method that is similar. In that method, the first step is*

$$u^{n+\frac{1}{2}}_j = u^n_j - \frac{\sigma}{2}\left(f^n_{j+\frac{1}{2},l} - f^n_{j-\frac{1}{2},r}\right) , \tag{C.5a}$$

and a second step of

$$u_{j+\frac{1}{2},l} = u^{n+\frac{1}{2}}_j + \frac{1}{2}\widetilde{\Delta_j u} , \tag{C.5b}$$

and

$$u_{j+\frac{1}{2},r} = u^{n+\frac{1}{2}}_{j+1} + \frac{1}{2}\widetilde{\Delta_{j+1} u} . \tag{C.5c}$$

C.2.3 Component-Wise Extension

A third approach is also available. This approach involves the separate limiting of the flux vector and the solution variable. It has been used by [200] with a high-order Lax-Friedrichs solver. This solver makes use of the identity, $f = au$, which implies that

$$\frac{\partial f}{\partial x} = a\frac{\partial u}{\partial x} , \tag{C.6}$$

which gives an equivalent form to that used above with a Lax-Wendroff approach. Specifically this can be written

$$u^{n+\frac{1}{2}}_{j+\frac{1}{2},l} = u_j + \frac{1}{2}\left(\widetilde{\Delta_j u} - \sigma\widetilde{\Delta_j f}\right)_j , \tag{C.7a}$$

and

$$u^{n+\frac{1}{2}}_{j+\frac{1}{2},r} = u_{j+1} - \frac{1}{2}\left(\Delta_{\widetilde{j+1}} u - \sigma\Delta_{\widetilde{j+1}} f\right) , \tag{C.7b}$$

where

$$\widetilde{\Delta_j f} = Q(i,r)\,\Delta_{j-\frac{1}{2}} f . \tag{C.7c}$$

Similar to the approach taken with the interpolation of the dependent variables, $r = \Delta_{j+\frac{1}{2}} f / \Delta_{j-\frac{1}{2}} f$ and $\Delta_{j-\frac{1}{2}} f = f_j - f_{j-1}$. Again for the scalar wave equation, this is equivalent to the Lax-Wendroff type of time differencing.

C.3 Method for Extension to Systems

This section concerns itself with the subject of extending the methods described in the previous section to systems of equations. I deal with the specific case of the Euler equations for the conservation of mass, momentum, and total energy.

The above system of equations can be written is a so-called primitive variable form. It has been suggested that this system of variable should be used to determine cell-edge states [234, 122]. In the above form the variables are conserved quantities $(\rho, m, E)^T$, but in the form given below the variables are $(\rho, u, e)^T$, the density, velocity, and internal energy. This follows the description of Roe's solver given in Appendix B. This set of equations is

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} = 0 , \tag{C.8a}$$

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + \frac{1}{\rho} \frac{\partial p}{\partial x} = 0 , \tag{C.8b}$$

and

$$\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} + \frac{p}{\rho} \frac{\partial u}{\partial x} = 0 . \tag{C.8c}$$

The equations in primitive form give a much simpler system than the Euler equations. The flux Jacobian is

$$A = \begin{bmatrix} u & \rho & 0 \\ \frac{e(\gamma-1)}{\rho} & u & \gamma - 1 \\ 0 & \frac{p}{\rho} & u \end{bmatrix} . \tag{C.9a}$$

Again, the eigenvalues of this matrix are

$$\left(\lambda^1, \lambda^2, \lambda^3\right) = (u - c, u, u + c) . \tag{C.9b}$$

The right eigenvectors form a matrix

$$R = \left(r^1, r^2, r^3\right) = \begin{bmatrix} 1 & 1 & 1 \\ -\frac{c}{\rho} & 0 & \frac{c}{\rho} \\ \frac{p}{\rho^2} & \frac{p}{(\gamma-1)\rho^2} & \frac{p}{\rho^2} \end{bmatrix} . \tag{C.9c}$$

and by using

$$z_1 = (\gamma - 1)\rho^2 ,$$

and

$$z_2 = 2\gamma p ,$$

the left eigenvectors form a matrix

$$R^{-1} = \begin{bmatrix} l^1 \\ l^2 \\ l^3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2\gamma} & -\frac{\rho}{2c} & \frac{z_1}{z_2} \\ \frac{\gamma}{\gamma - 1} & 0 & -\frac{2z_1}{z_2} \\ \frac{1}{2\gamma} & \frac{\rho}{2c} & \frac{z_1}{z_2} \end{bmatrix} . \tag{C.9d}$$

Of the methods available for extending the scheme outlined in the previous section, the characteristic decomposition due to Roe [53] is the most common. In this method, a similarity transform takes the variable from the conservative form to a characteristic form. Each variable can then be computed at the cell edges from its characteristic contributions. This methodology can also be applied to the primitive variables in a similar manner. The basic theory of Roe's method is given in Appendix B.

Thus, each characteristic is limited separately in defining the new cell-edge value of **U**. For this purpose, I define

$$\widetilde{\Delta_j u} = \sum_{k=1}^{3} r^k \widetilde{\Delta_j \alpha}^k , \tag{C.10}$$

where

$$\widetilde{\Delta_j \alpha} = Q(1, r)\, \Delta_{j-\frac{1}{2}} \alpha \tag{C.11}$$

for each component of **U** where $r = \Delta_{j+\frac{1}{2}} \alpha / \Delta_{j-\frac{1}{2}} \alpha$.

The characteristic approach must also be integrated into the attainment of temporal accuracy. Each wave in the above decomposition travels at different speeds and they can also travel in different directions. For this reason, the cell-edge quantities are computed from the following formulas:

$$\mathbf{U}_{j+\frac{1}{2},l} = \mathbf{U}_j + \frac{1}{2} \sum_{k=1}^{3} r^k \left(1 - \eta^k\right) \widetilde{\Delta_j \alpha}^k , \tag{C.12a}$$

and

$$\mathbf{U}_{j+\frac{1}{2},r} = \mathbf{U}_{j+1} - \frac{1}{2} \sum_{k=1}^{3} r^k \left(1 + \eta^k\right) \widetilde{\Delta_{j+1} \alpha}^k , \tag{C.12b}$$

here $\eta^k = \lambda^k \sigma$. Colella [234] reports a more robust characteritic decomposition that is described and tested in Appendix D.

This method is aesthetically pleasing because the coupled nonlinear system is

locally reduced to a set of decoupled scalar equations. Because of this, the theory developed and applied to simpler model problems carries over without interference to systems. On the other hand, the expense associated with procedure (especially when multidimensional or more complex systems are considered) makes them less attractive than other alternatives. A modification of this method that is touted as increasing the robustness of the reconstruction is given in [234]. This method takes into account the direction of wave carrying information and only allows physically meaningful reconstructions to occur.

The other options described in Section C.2 are somewhat more straightforward to implement for systems of equations. The two-step method is simply applied in a vector fashion, i.e.,

$$\mathbf{U}^{n+\frac{1}{2}}_{j+\frac{1}{2},l} = \mathbf{U}^{n}_{j+\frac{1}{2},l} - \frac{\sigma}{2}\left[\mathbf{F}\left(\mathbf{U}^{n}_{j+\frac{1}{2},l}\right) - \mathbf{F}\left(\mathbf{U}^{n}_{j-\frac{1}{2},r}\right)\right] . \tag{C.13a}$$

and

$$\mathbf{U}^{n+\frac{1}{2}}_{j+\frac{1}{2},r} = \mathbf{U}^{n}_{j+\frac{1}{2},r} - \frac{\sigma}{2}\left[\mathbf{F}\left(\mathbf{U}^{n}_{j+\frac{1}{2},l}\right) - \mathbf{F}\left(\mathbf{U}^{n}_{j+\frac{1}{2},r}\right)\right] . \tag{C.13b}$$

Similarly the component-wise extension method can be extended by using limited values of the flux function for each of a system's equations. Thus, the method can be written

$$\mathbf{U}^{n+\frac{1}{2}}_{j+\frac{1}{2},l} = \mathbf{U}_j + \frac{1}{2}\left(\widetilde{\Delta_j \mathbf{u}} - \widetilde{\Delta_j \mathbf{f}}\right) , \tag{C.14a}$$

and

$$\mathbf{U}^{n+\frac{1}{2}}_{j+\frac{1}{2},r} = \mathbf{U}_{j+1} - \frac{1}{2}\left(\widetilde{\Delta_{j+1} \mathbf{u}} - \widetilde{\Delta_{j+1} \mathbf{f}}\right) . \tag{C.14b}$$

For both of these methods, the computation of the cell-edge value could be done in either conservative, primitive, or characteristic variables. The advantage of the two-step or the component-wise extension methods can only be obtained if the interpolation is done in either the conservative or primitive variables because of the relative simplicity of each formulation.

Another issue of some importance is the application of limiters in computing the piecewise polynomials. It is common practice to use a compressive limiter such as superbee on the field that produces the contact discontinuity. The compression given by the limiter maintains the sharpness of the interface. The same limiter when applied to shocks or rarefactions can produce entropy violating solutions. For the characteristic decomposition the implementation of this is quite clear. For other methods not involving characteristic decomposition it is usual practice to apply the compressive limiter to the computation of the density profile [122].

Table C.1: Abbreviations for the methods used in this study.

Scheme	Abbreviation
Characteristic-conservative variables	CC
Characteristic-primitive variables	PC
Two-Step-conservative variables	CR
Two-Step-primitive variables	PR
Component-wise-conservative variables	CF
Component-wise-primitive variables	PF

C.4 Comparison of Methods

In the following section, I compare the performance of the methods for several standard test problems for the Euler equations in one space dimension. The results of this discussion should provide guidance for more complex systems of equations as well as guidance in a route to take in extending these methods to multidimensional problems. Table C.1 list the abbreviations used in this section to describe the methods.

C.4.1 Sod's Problem

The solutions to Sod's problem can be seen in Figs. C.1-C.6. In general, the solutions are quite good and exhibit the qualities one would expect with a high-resolution numerical solution.

The solutions found with the CC method are seen in Fig. C.1. They are qualitatively quite good, with the only problem being the glitch in the velocity at the end of the rarefaction wave. With the PC method the velocity glitch is gone, but a small rise is before to the shock. As can be seen in Fig. C.2, the density profile is nearly identical to that found with the CC method.

With the two-step formulation, the solutions are again quite good as can be seen in Figs. C.3 and C.4. The major problems can be seen with the velocity profiles where small problems exist with at the end of the rarefaction wave and in the post shock region of the flow. These problems are not major in nature. Major features of the flow field such as the shock, contact discontinuity and rarefaction wave are resolved well.

The component-wise extension of the schemes has a few more problems. In Figs. C.5 and C.6 the solutions are shown. The shock wave is exceptionally sharp, improved over the other methods, but in both the conservative and primitive variable formulation there are a number of small oscillations in the velocity solution between

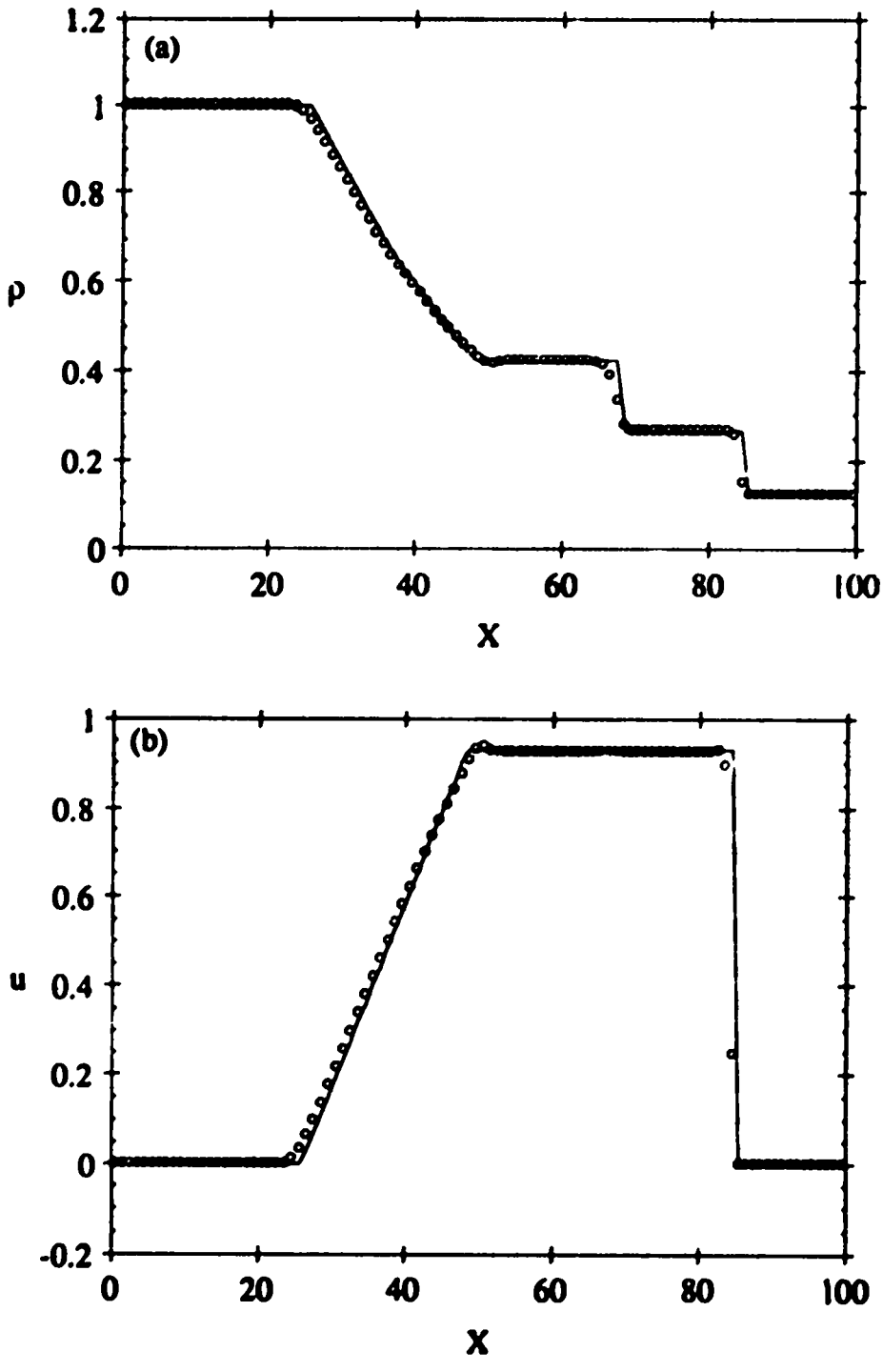


Figure C.1: Sod's problem computed with the characteristic formulation with conservative variables. In these figures, the solid line denotes the exact solution, whereas the circles denote the approximate numerical solution.

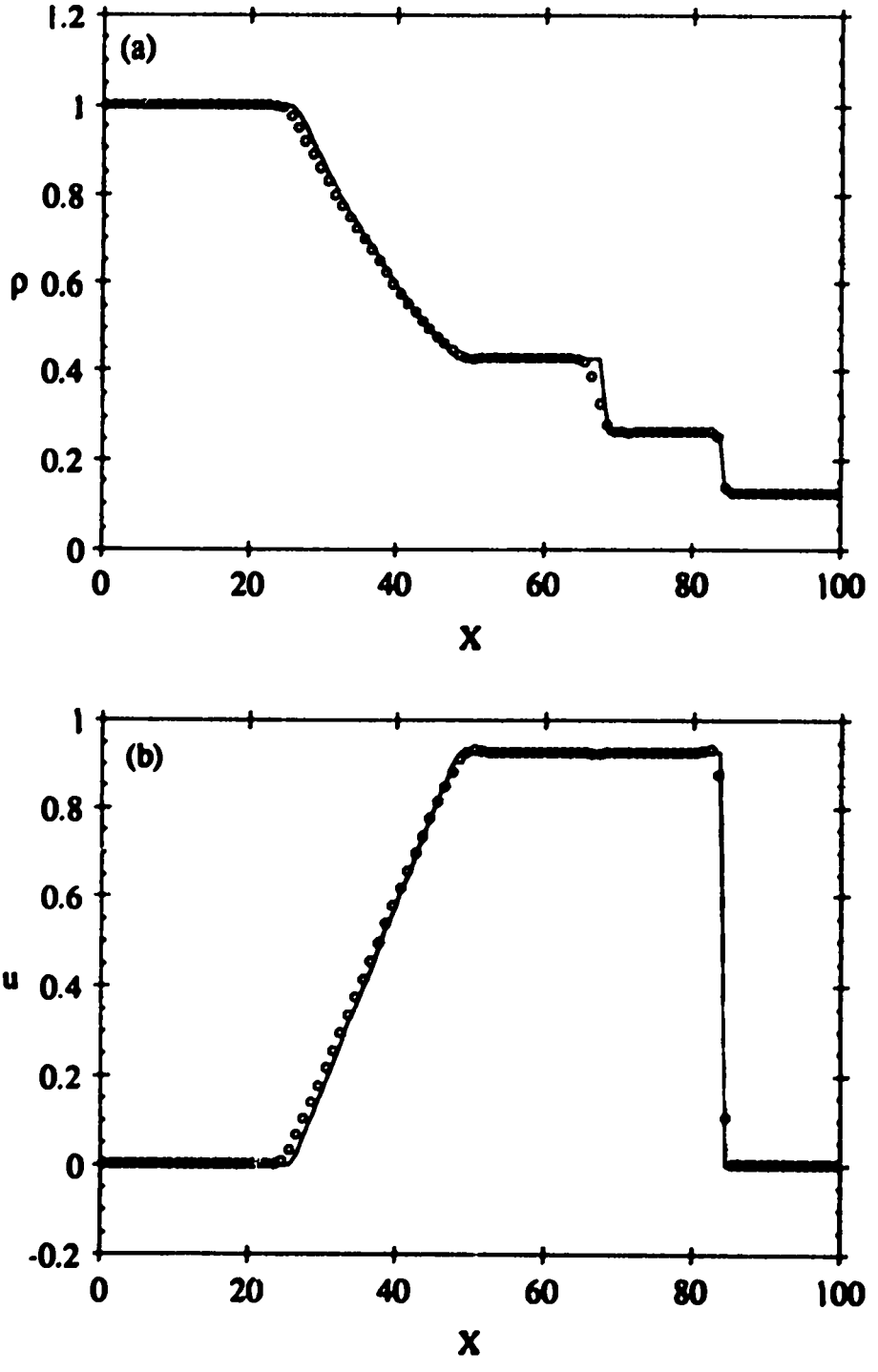


Figure C.2: Sod's problem computed with the characteristic formulation with primitive variables.

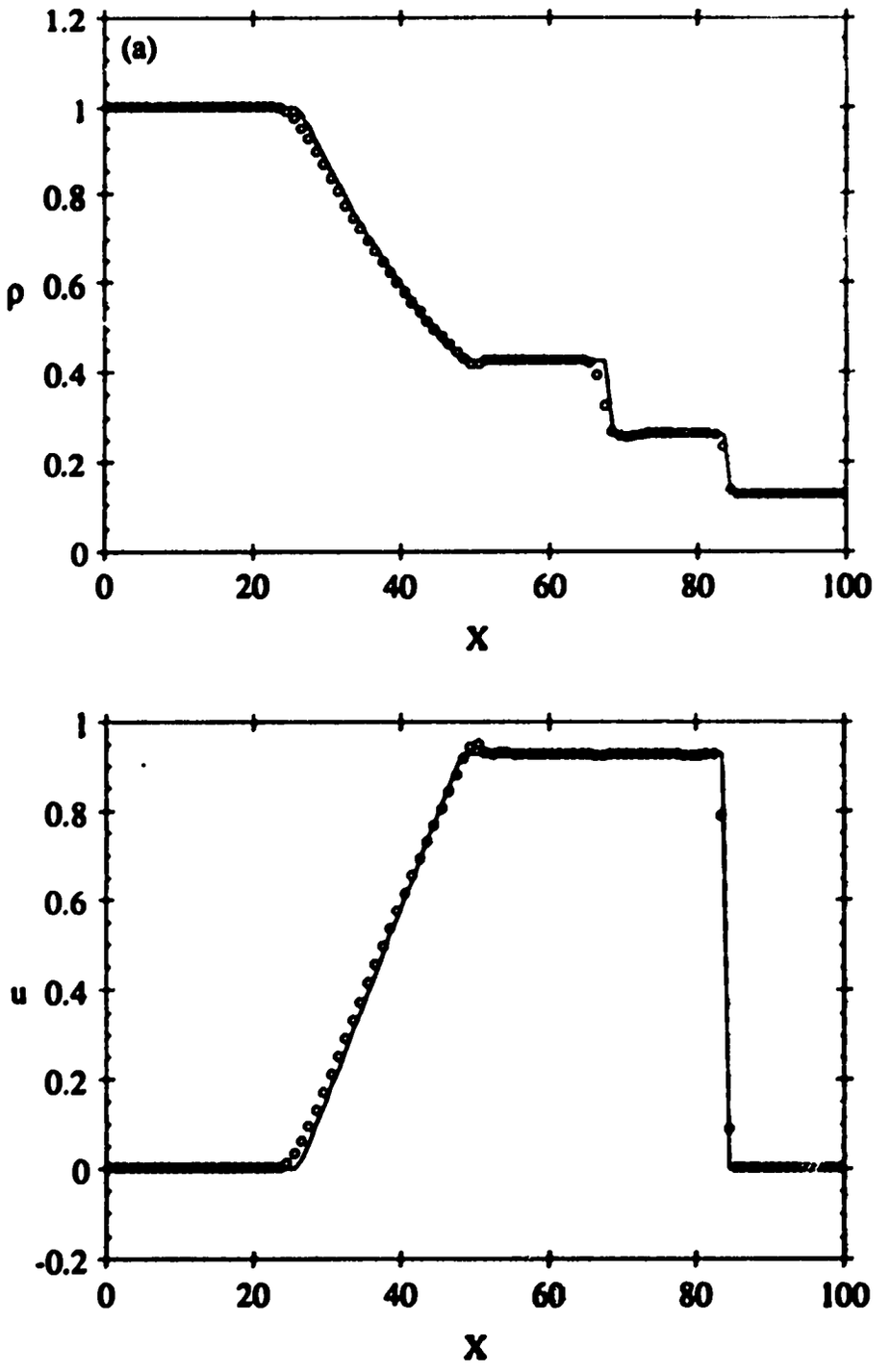


Figure C.3: Sod's problem computed with the two-step formulation with conservative variables.

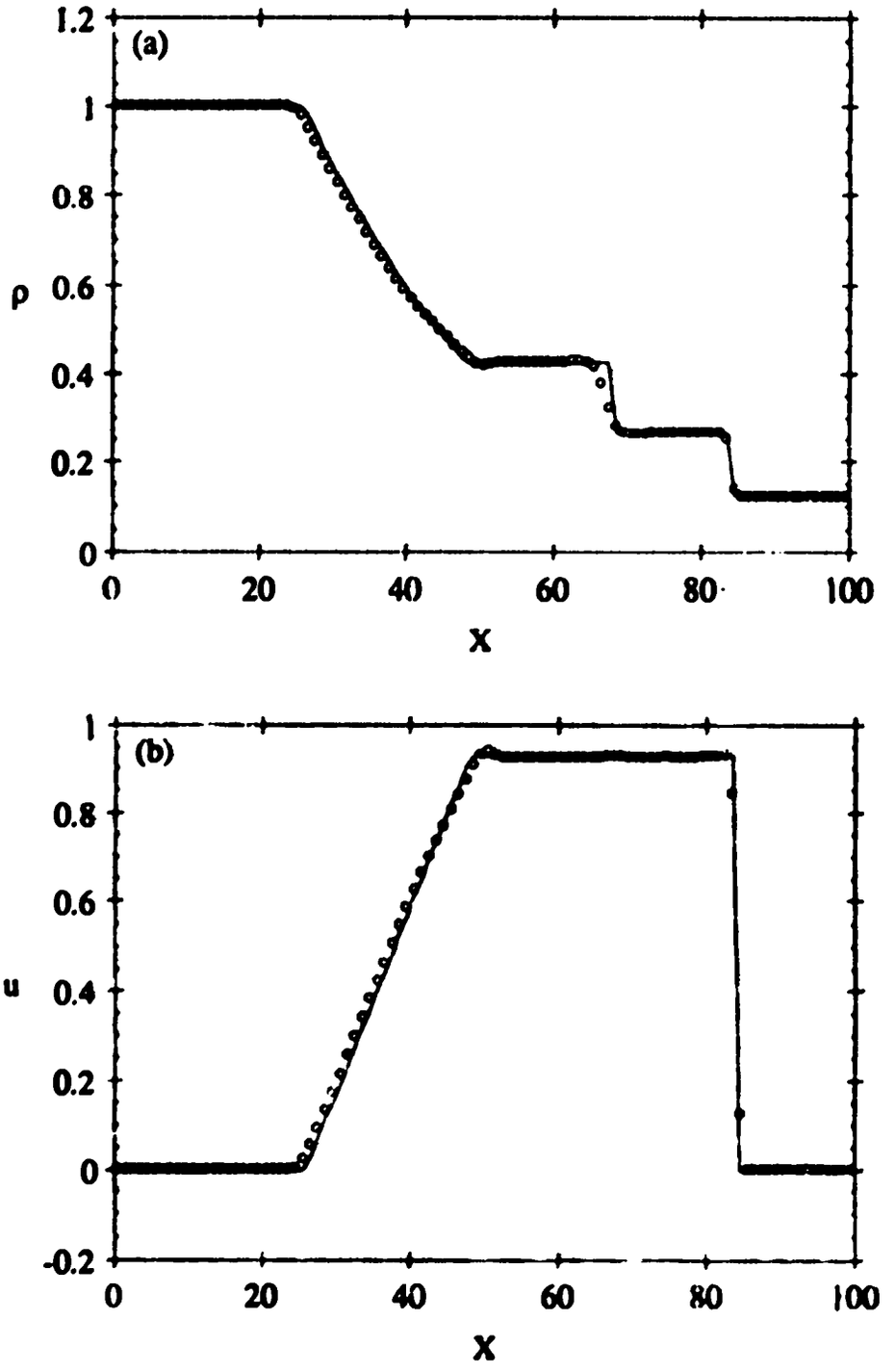


Figure C.4: Sod's problem computed with the two-step formulation with primitive variables. Note the small spikes at the end of the rarefaction waves and the post-shock spike in the velocity solution.

Table C.2: The L_1 error norms for each scheme on Sod's problem

Scheme	Density	Velocity
CC	5.86×10^{-3}	1.19×10^{-2}
PC	4.90×10^{-3}	6.14×10^{-3}
CR	5.26×10^{-3}	7.27×10^{-3}
PR	5.45×10^{-3}	7.58×10^{-3}
CF	5.34×10^{-3}	9.33×10^{-3}
PF	6.20×10^{-3}	1.22×10^{-2}

the rarefaction and shock waves. In this case, these oscillations are not destructive, but detract from the overall quality of the solution.

In Table C.2, the L_1 norm errors using these methods are shown. In these terms the best solution is the PC method with both of the two step methods of slightly lower quality. The PF method is the worst, with the CC formulation slightly better. However, the better qualitative appearance of the CC makes it much superior to the PF method.

C.4.2 Lax's Problem

The solutions to this problem by the methods discussed in this appendix are shown in Figs. C.7-C.12. Again the solutions are quite good across the board, but problems with the methods show more strongly in the density profiles. The region between the shock wave and the contact discontinuity is sensitive to the limiter used, and in the non characteristic methods, problems show up.

Figures C.7 and C.8 show the CC and PC solutions to Lax's problem, respectively. The only problem with these solutions is evident in the PC velocity solution where a small dip in the velocity is present coincident with the contact discontinuity. This is an artifact of the compressive superbee limiter used on the linearly degenerate wave.

Figures C.9-C.12 show the solutions found with other methods. These solutions all share common characteristics. The contact discontinuity causes oscillations in the solutions as evident in both the density and velocity profiles. These oscillations are more severe in the primitive variable formulations. These oscillations can be controlled through another choice of a limiter to apply to the density interpolation.

In terms of L_1 error (see Table C.3), the conclusions that are drawn are somewhat different to those found with Sod's problem. The velocity errors are very close in magnitude and no real conclusions can be drawn from them. The density errors

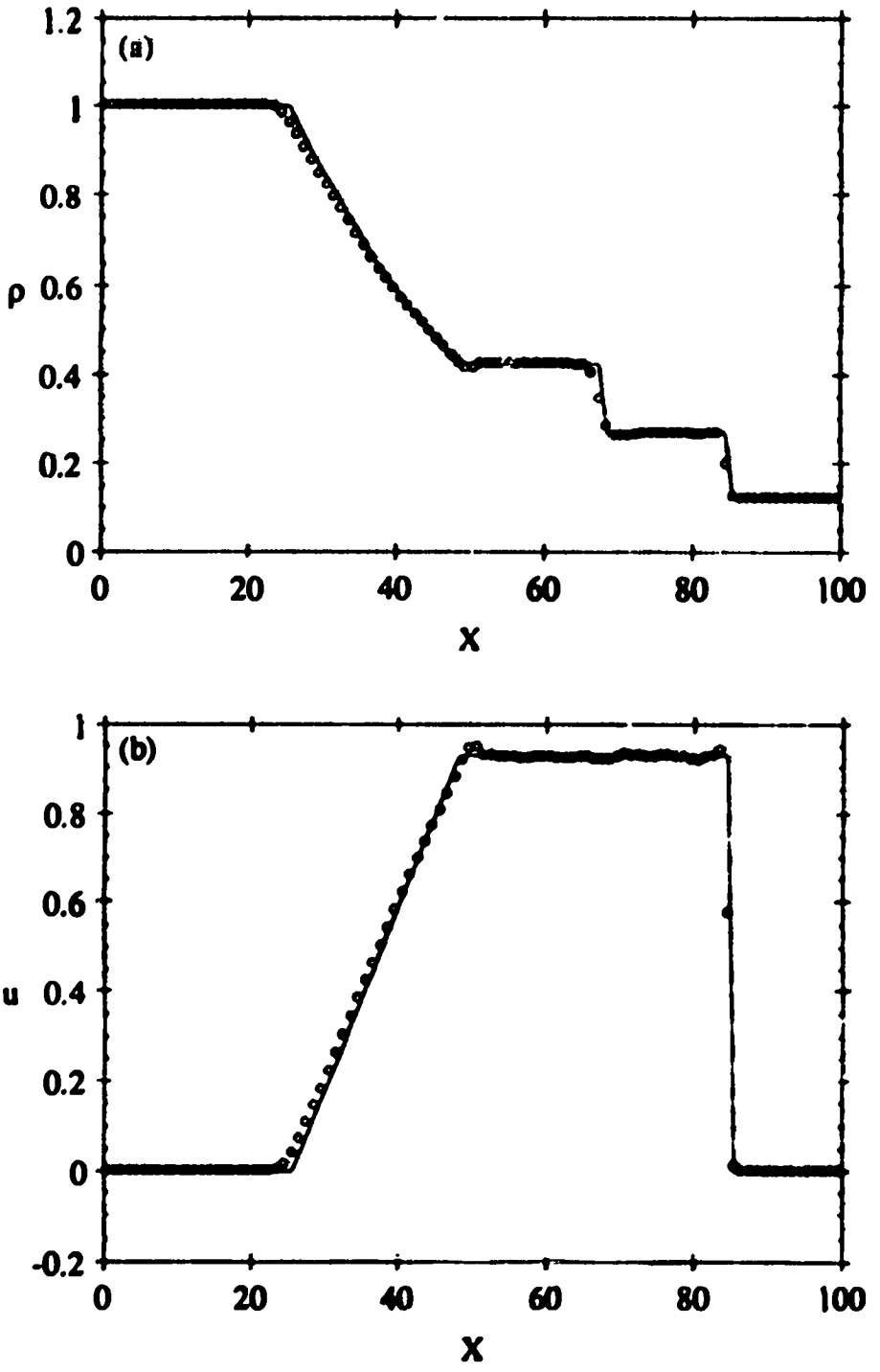


Figure C.5: Sod's problem computed with the component-wise formulation with conservative variables. Note the small oscillations in the velocity solution between the rarefaction and shock waves.

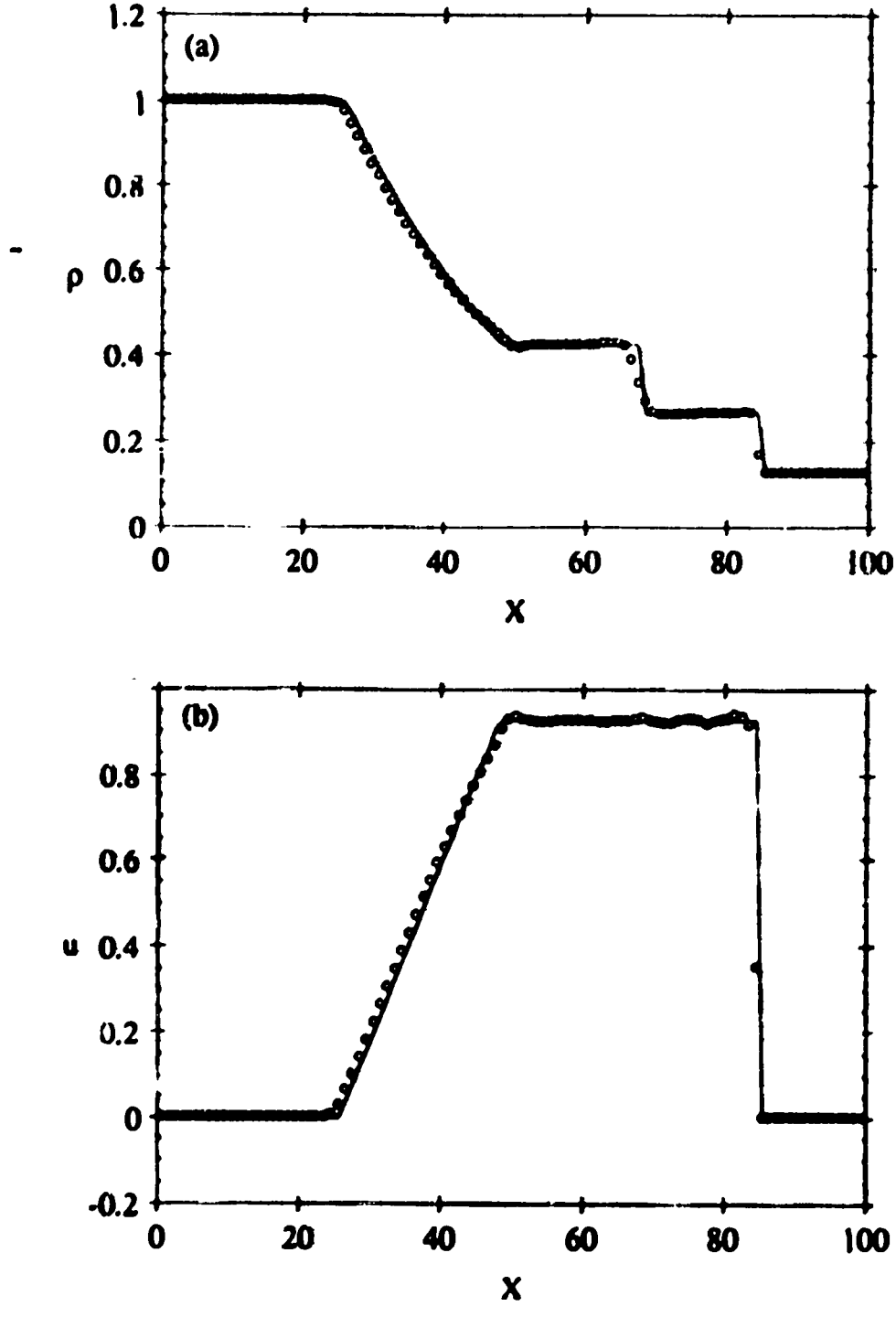


Figure C.6: Sod's problem computed with the component-wise formulation with primitive variables. Note the small oscillations in the velocity solution between the rarefaction and shock waves.

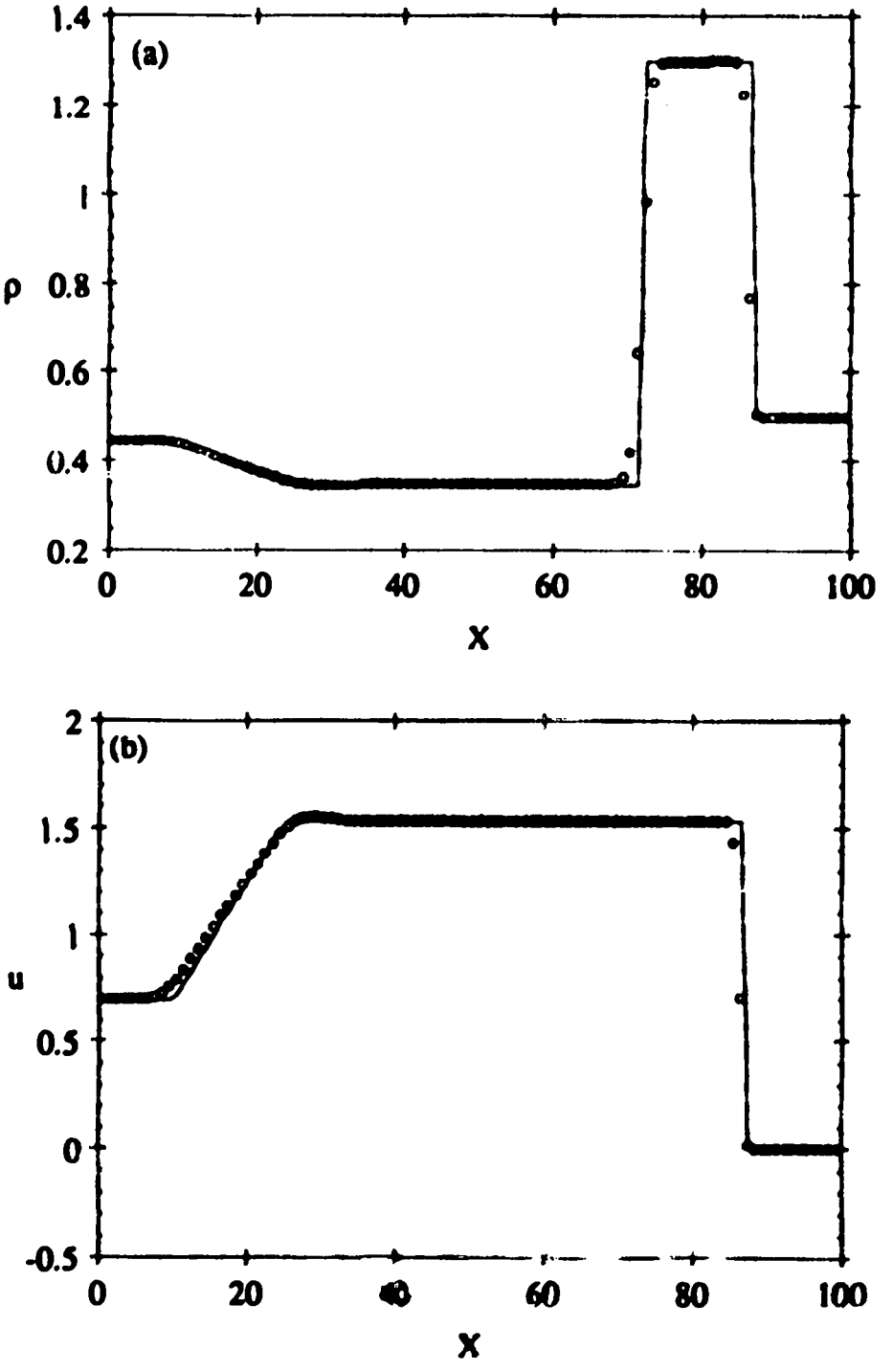


Figure C.7: Lax's problem computed with the characteristic formulation with conservative variables. With the exception of this solution, all the solutions to Lax's problem have small spikes or oscillations associated with the contact discontinuity. This is indicative of the overcompressive nature of the limiter placed on the density. The conservative characteristic formulation guards against this problem.

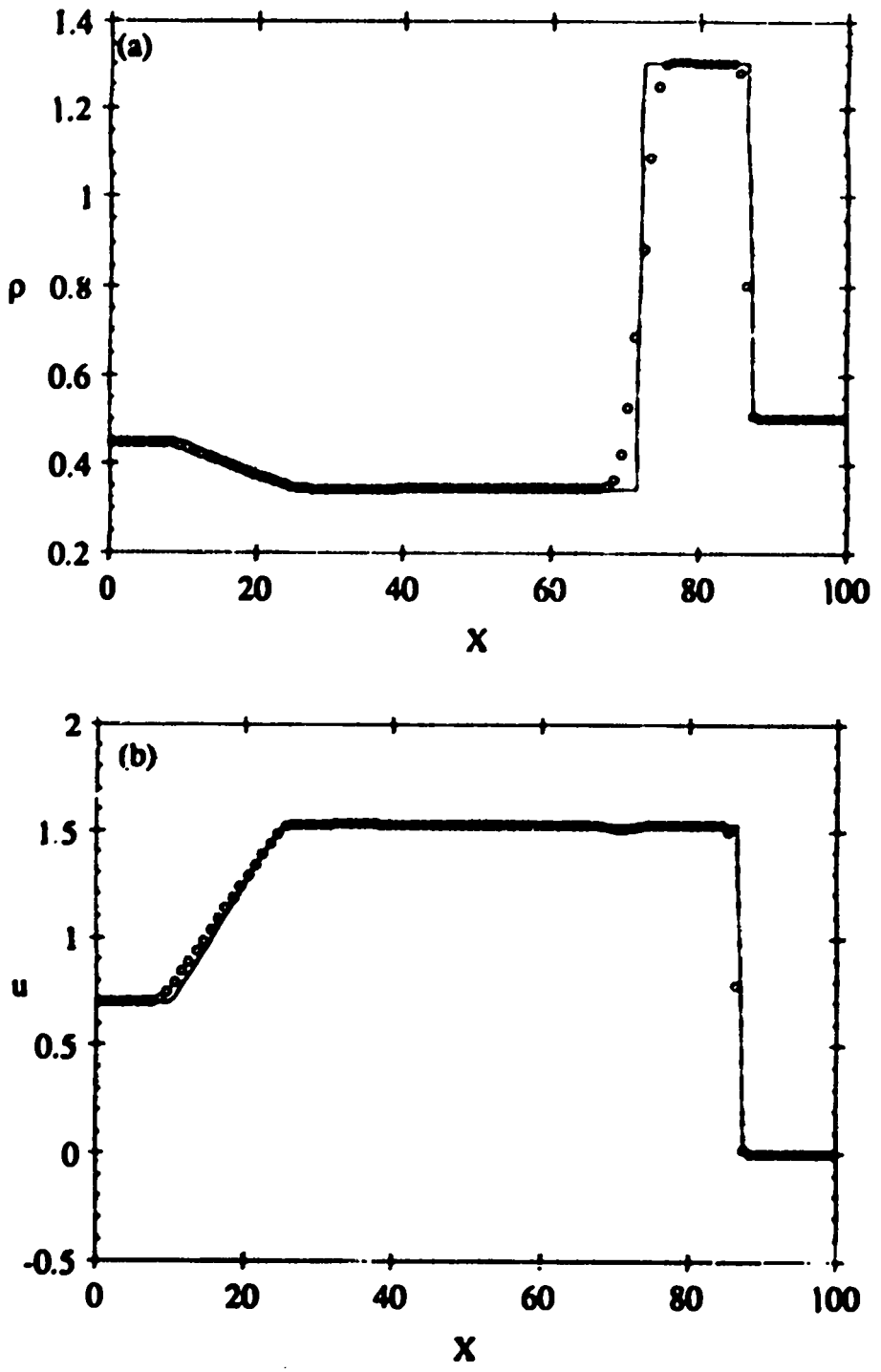


Figure C.8: Lax's problem computed with the characteristic formulation with primitive variables. Despite using a characteristic formulation, a small oscillation is present with the contact discontinuity.

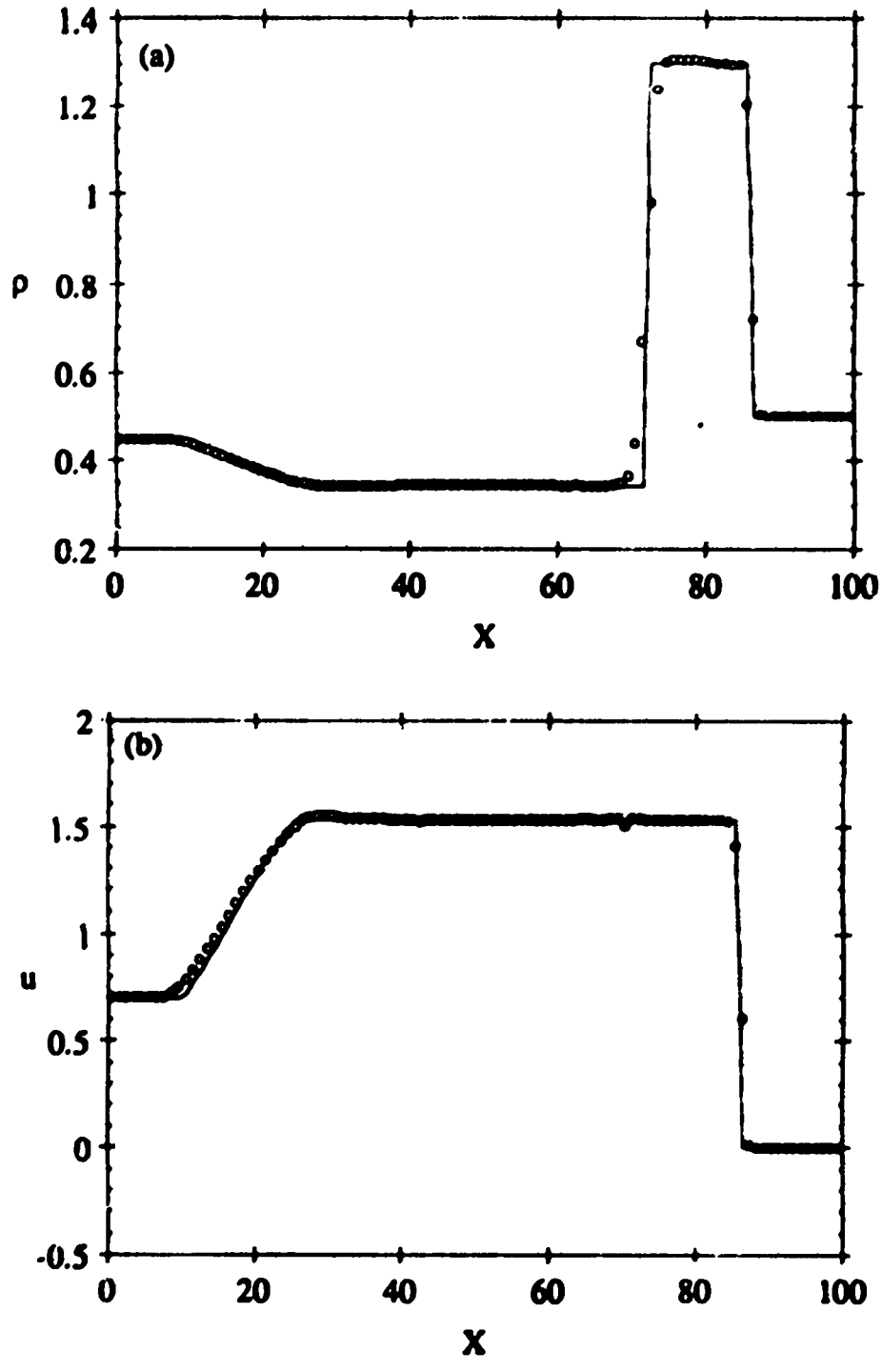


Figure C.9: Lax's problem computed with the two-step formulation with conservative variables.

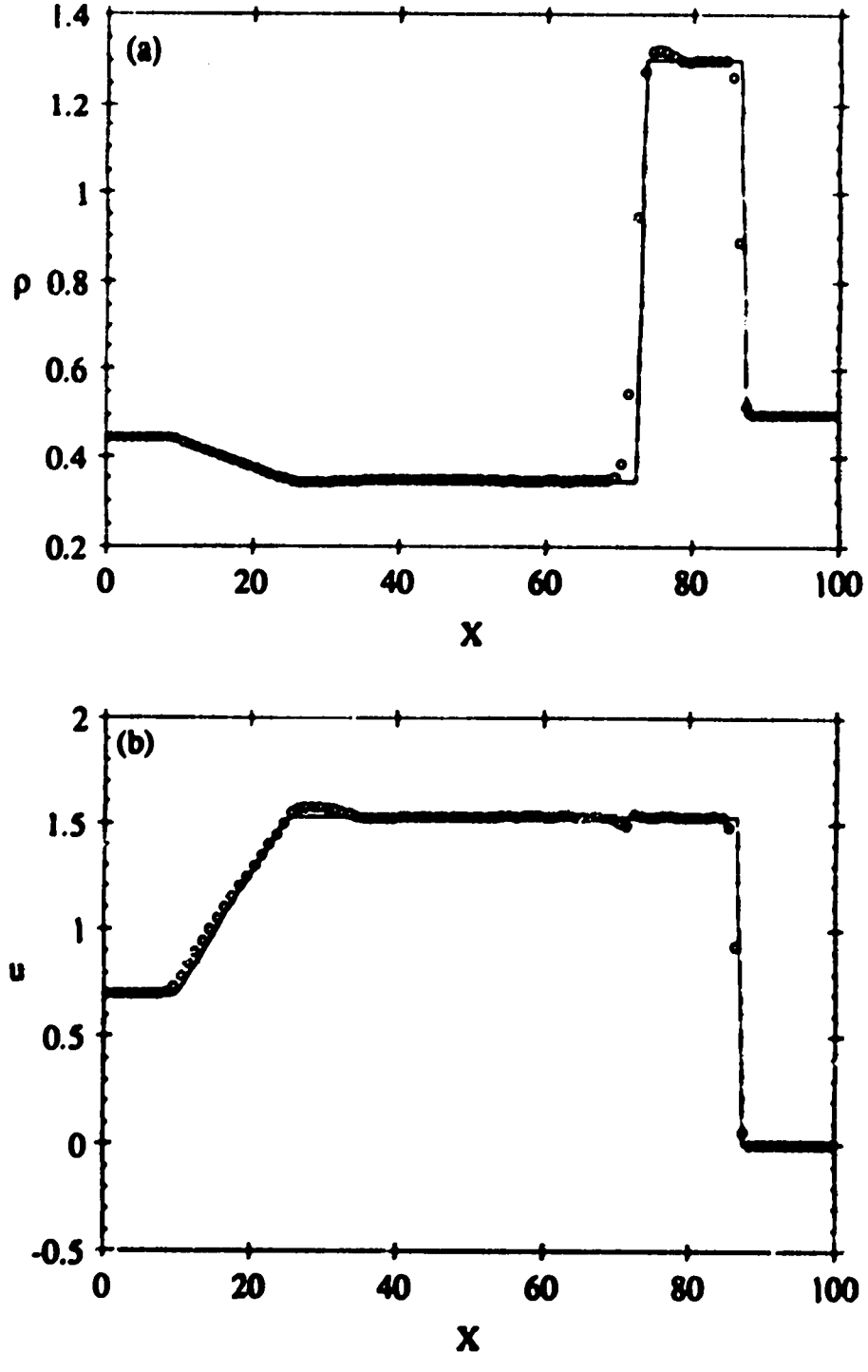


Figure C.10: Lax's problem computed with the two-step formulation with primitive variables.

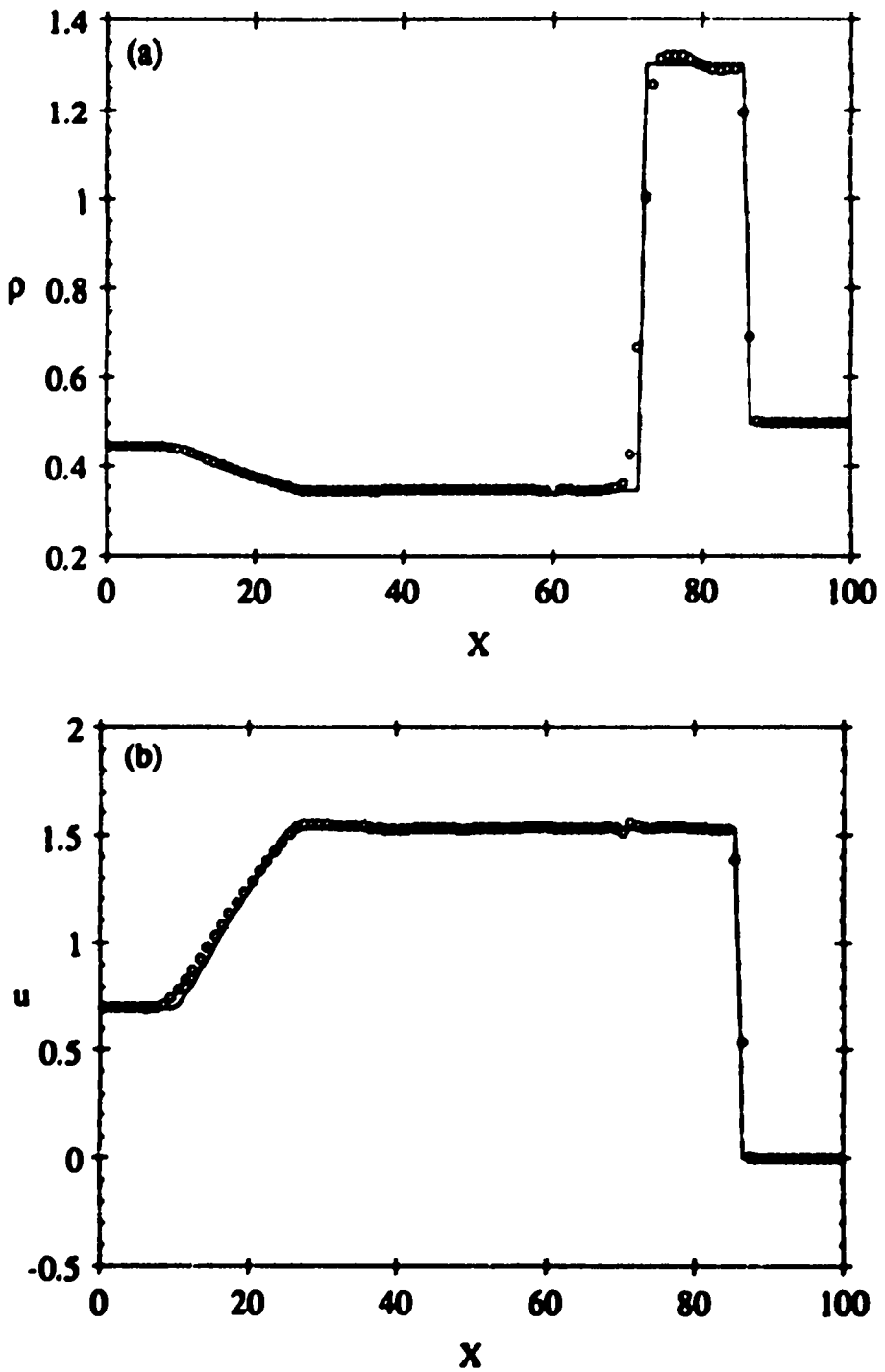


Figure C.11: Lax's problem computed with the component-wise formulation with conservative variables.

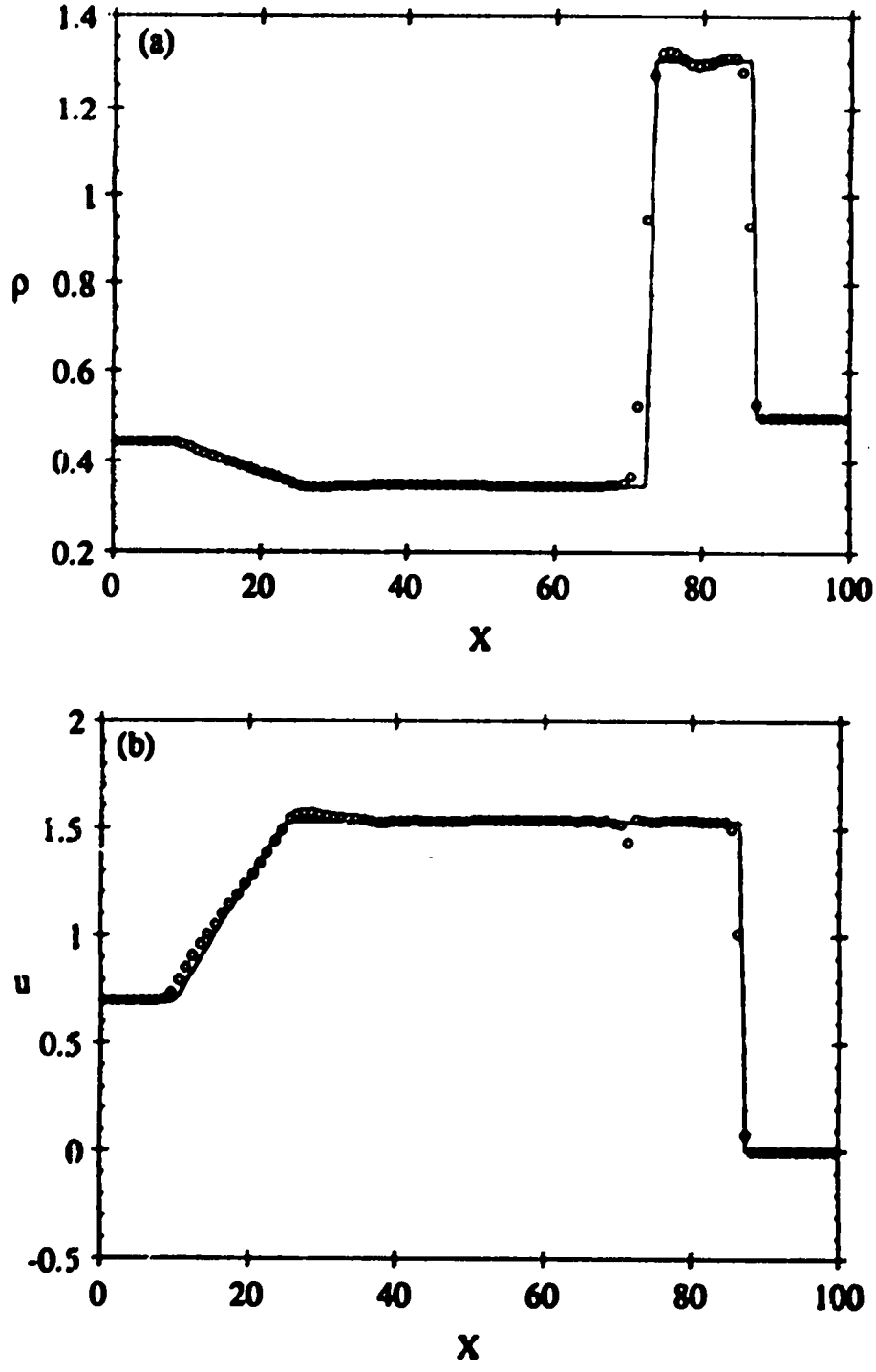


Figure C.12: Lax's problem computed with the component-wise formulation with conservative variables.

Table C.3: The L_1 error norms for each scheme on Lax's problem

Scheme	Density	Velocity
CC	1.46×10^{-2}	1.61×10^{-2}
PC	1.92×10^{-2}	1.42×10^{-2}
CR	1.30×10^{-2}	1.53×10^{-2}
PR	1.52×10^{-2}	1.61×10^{-2}
CF	1.29×10^{-2}	1.54×10^{-2}
PF	1.44×10^{-2}	1.62×10^{-2}

seem to favor the conservative formulations, but for the two-step or component-wise formulations the differences are not profound.

C.4.3 Vacuum Problem

As noted in Section C.2, one case in this study does not use Roe's approximate Riemann solver. The case of the vacuum problem considered below cannot use Roe's solver as explained in [231]. For this case, a more diffusive scheme is used to maintain physical solutions. This is the HLLE Riemann solver [30, 231, 128] (see Appendix B).

This method has several desirable properties: its simplicity, ease of implementation, and satisfaction of entropy inequalities. Reference [231] makes the suggestion for the computation of $b^r_{l,r}$ and $b^l_{l,r}$. The formulas are

$$b^r_{l,r} = \max\left(a_{r,max}, a_{l,r,max}\right) , \tag{C.15a}$$

and

$$b^l_{l,r} = \min\left(a_{l,min}, a_{l,r,min}\right) , \tag{C.15b}$$

where max and min refer to the maximum and minimum characteristic speeds at the respective locations. The values for $a_{l,r}$ come from the Roe linearization that is discussed below.

The solutions found with the CC, PC, PR, and PF (Figs. C.13, C.14, C.16 and C.18) methods are not worth much discussion. All of them are quite good and appear to be nearly identical in terms of resolution. Table C.4 shows this as well.

The solutions found with the CR and CF methods do warrant some discussion. The CR solution is shown in Fig. C.15 and the CF solution in Fig. C.17. Both solutions are of exceedingly poor quality. In fact if measure had not been taken to prevent this, the computer code should have blown up early in the solution process.

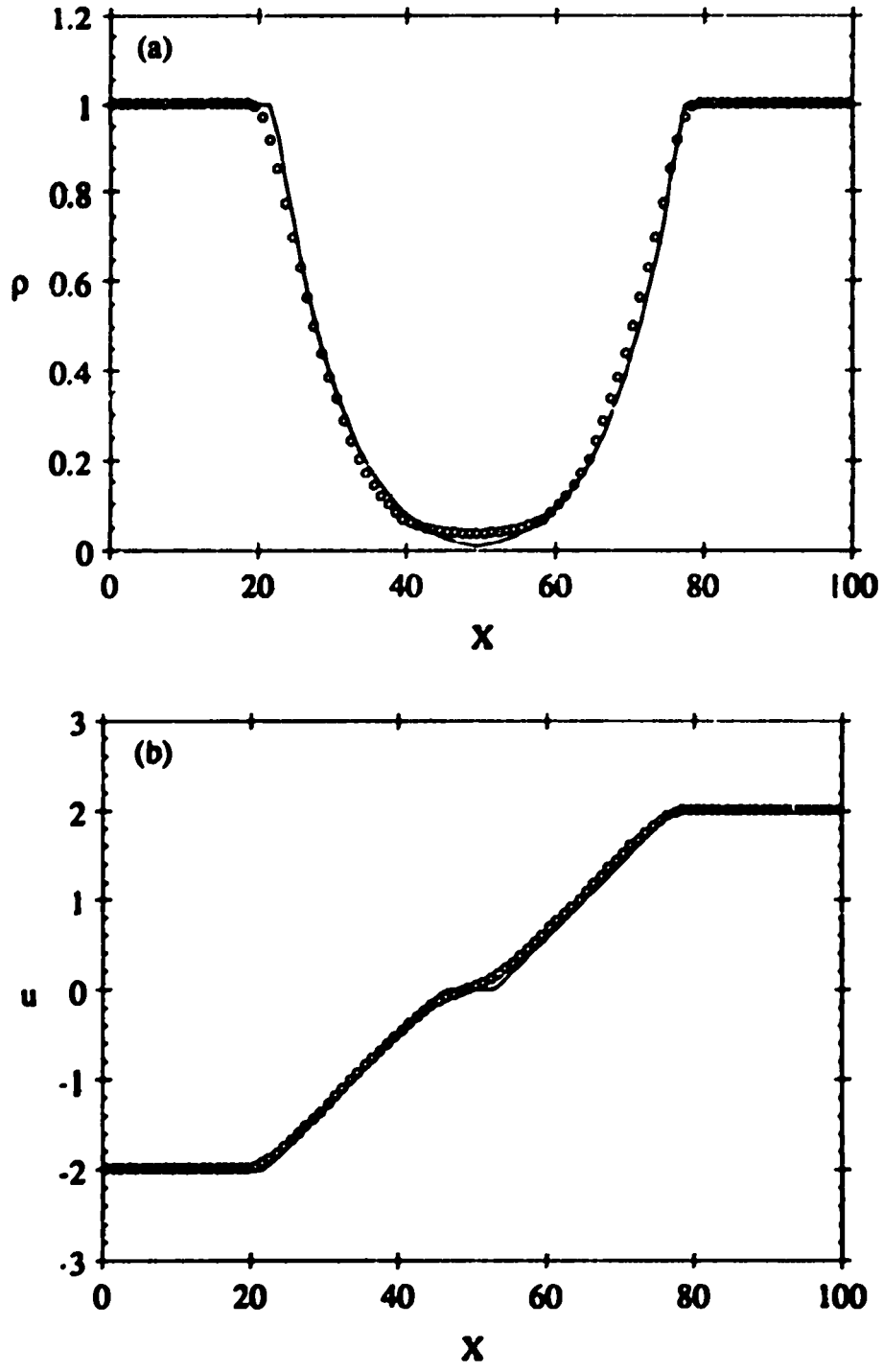


Figure C.13: The vacuum problem computed with the characteristic formulation with conservative variables.

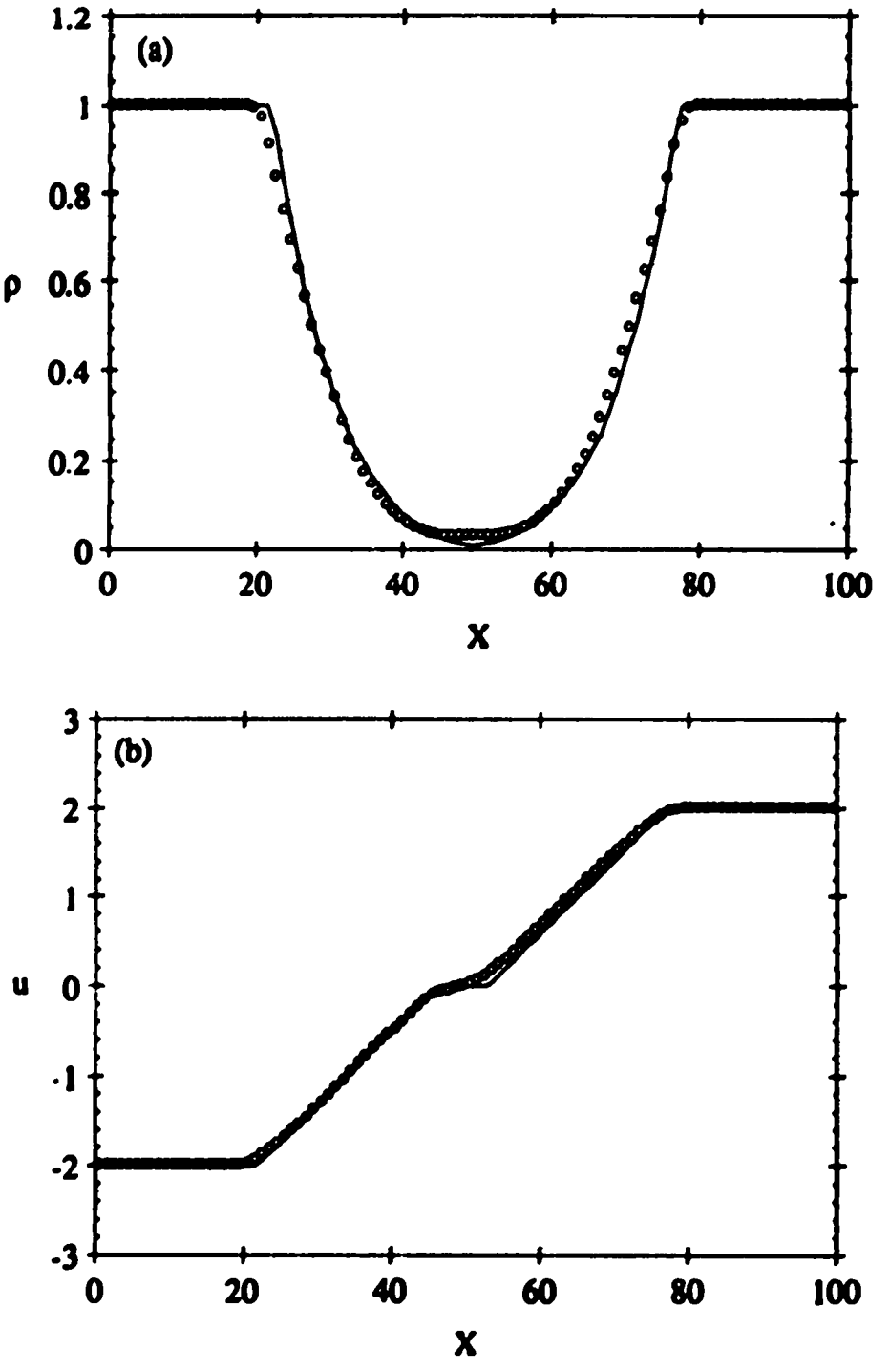


Figure C.14: The vacuum problem computed with the characteristic formulation with primitive variables.

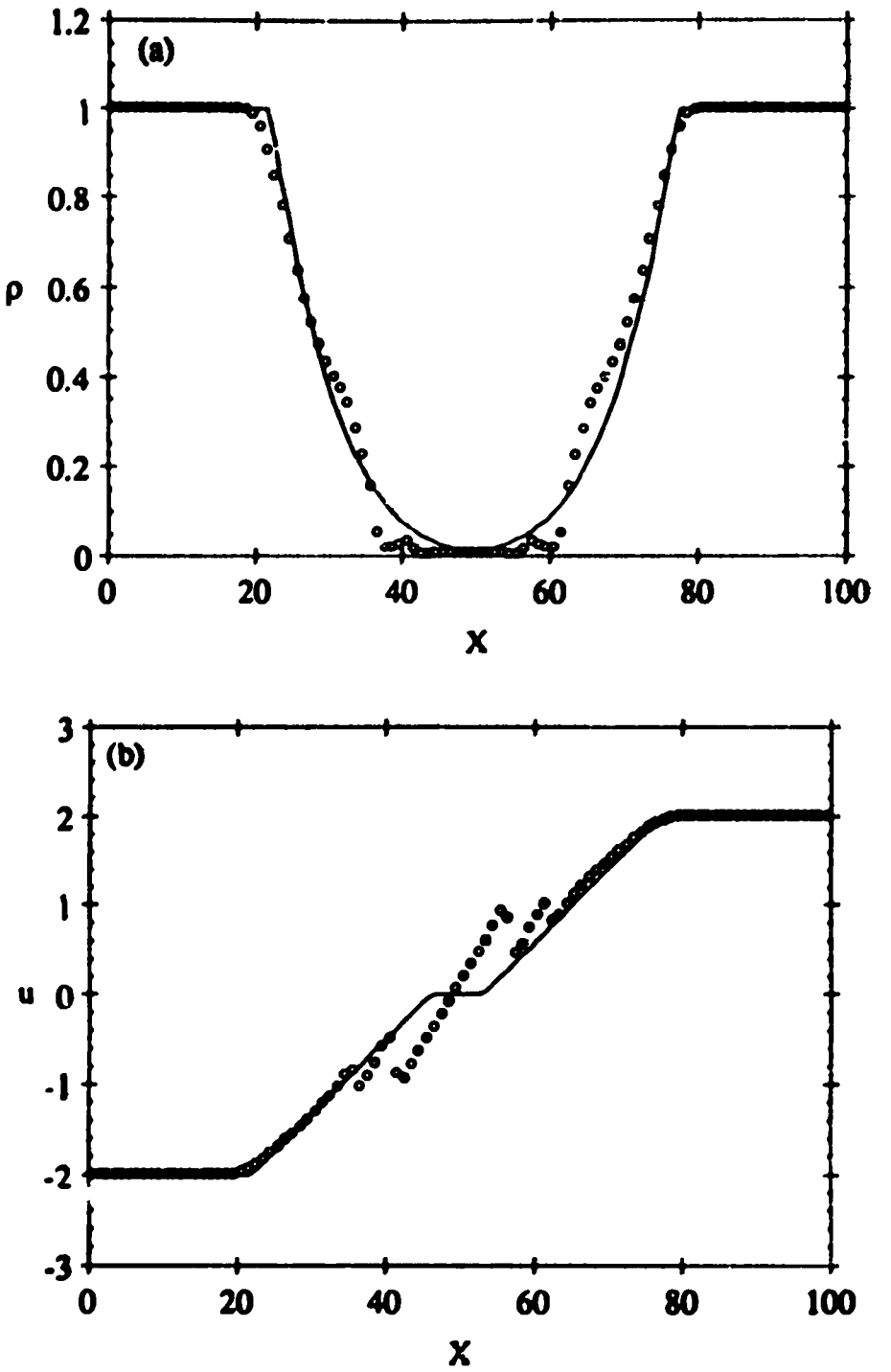


Figure C.15: The vacuum problem computed with the two-step formulation with conservative variables. The use of conservative variables with this flow is disastrous. The total energy has become negative in the region around $X = 50$.

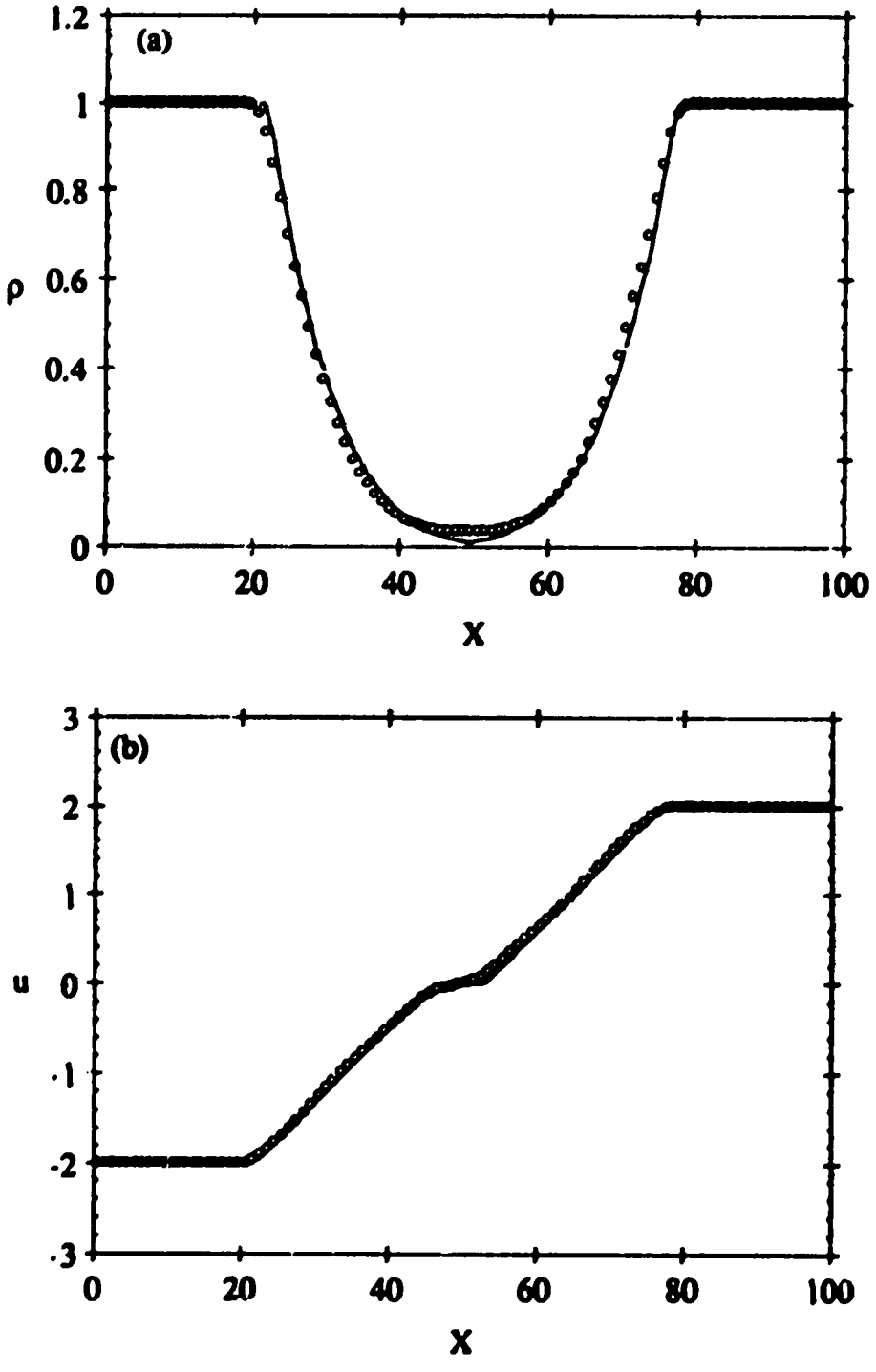


Figure C.16: The vacuum problem computed with the two-step formulation with primitive variables.

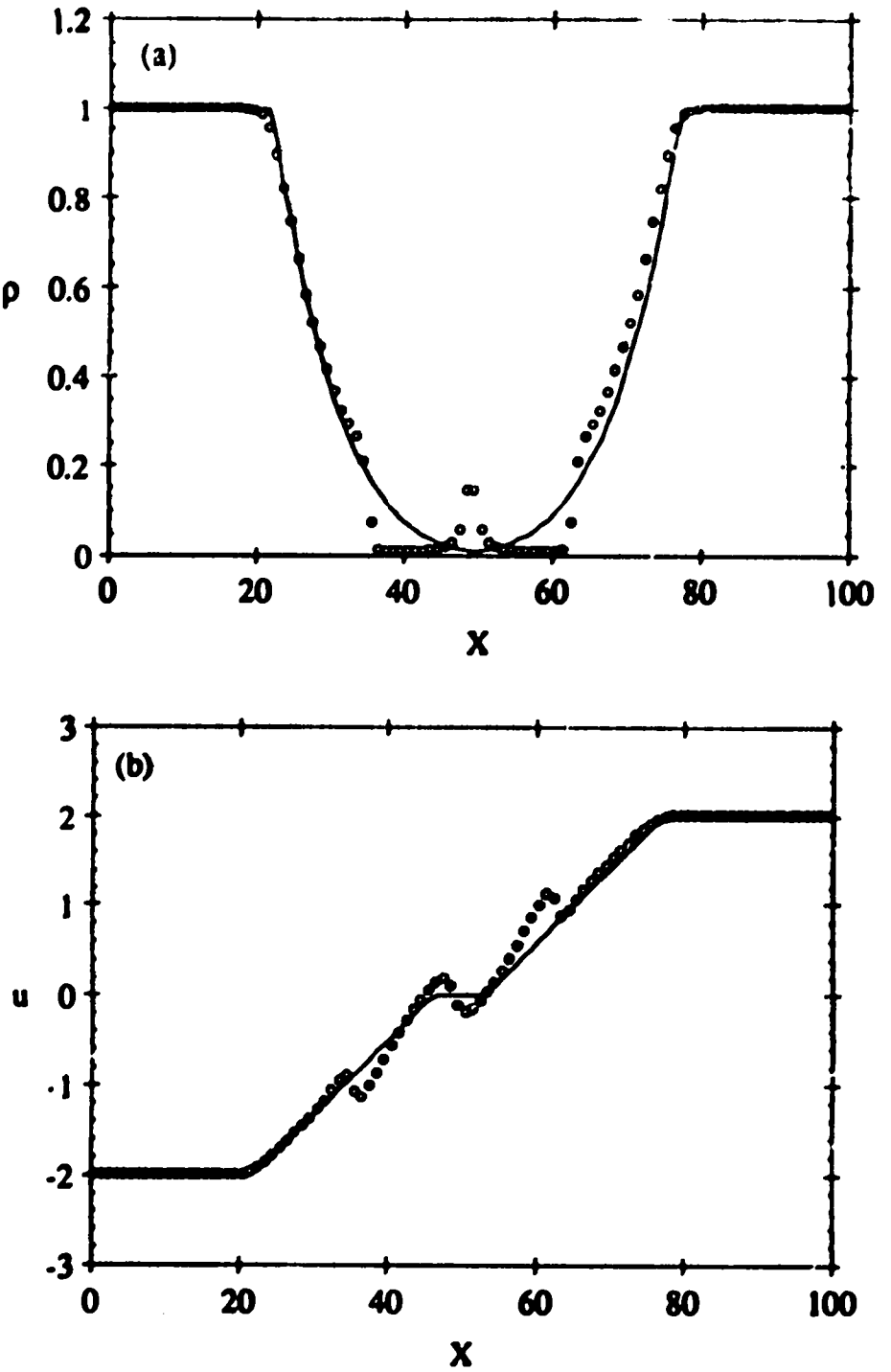


Figure C.17: The vacuum problem computed with the component-wise formulation with conservative variables. The conservative variables have not guaranteed that positive definite quantities (total energy) stay positive definite.

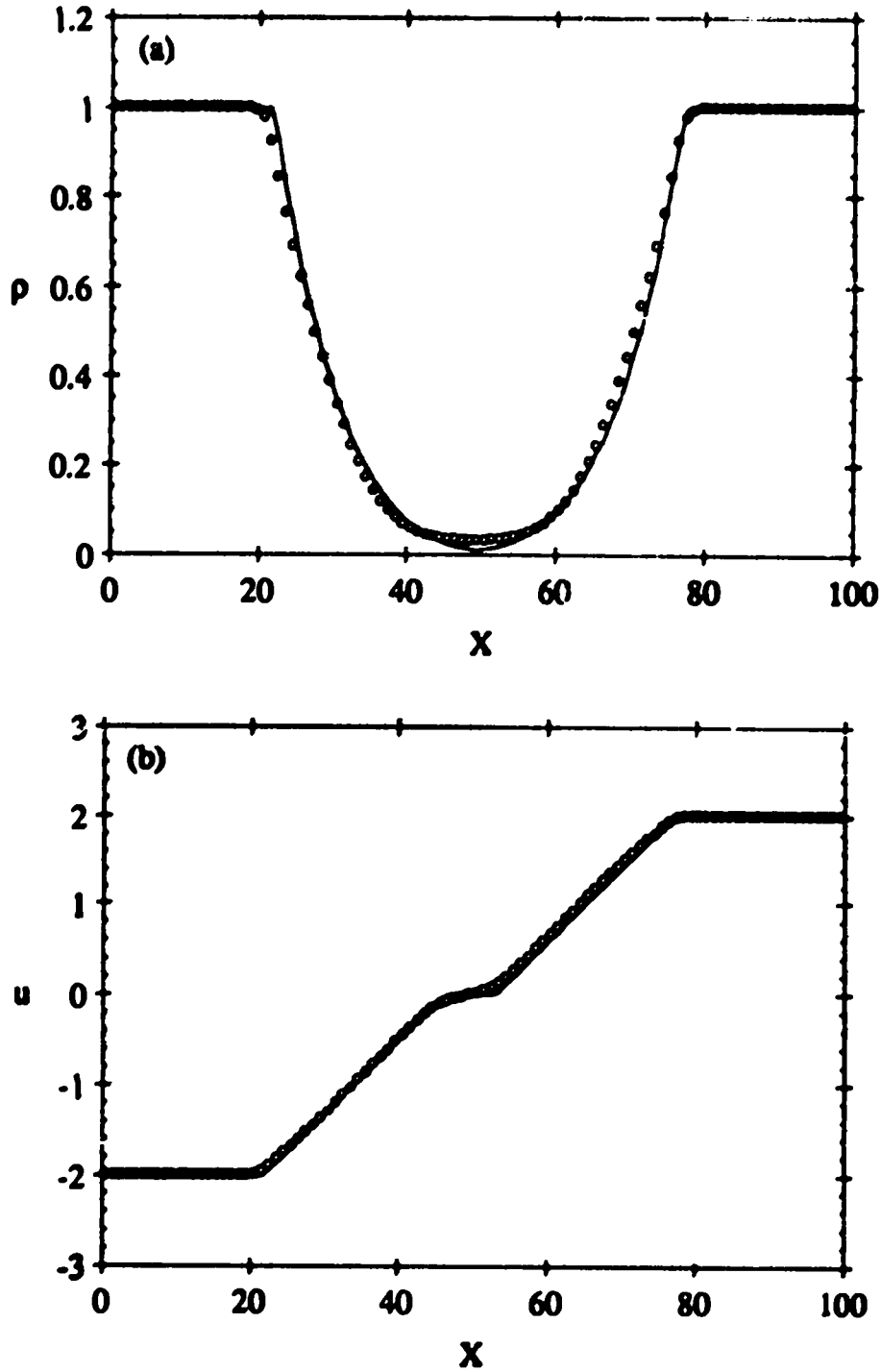


Figure C.18: The vacuum problem computed with the component-wise formulation with conservative variables.

Table C.4: The L_1 error norms for each scheme on the Vacuum problem

Scheme	Density	Velocity
CC	1.27×10^{-2}	2.63×10^{-2}
PC	1.24×10^{-2}	2.85×10^{-2}
CR	2.72×10^{-2}	1.00×10^{-1}
PR	1.20×10^{-2}	2.39×10^{-2}
CF	2.81×10^{-2}	5.85×10^{-2}
PF	1.20×10^{-2}	2.40×10^{-2}

This is because the total energy in the solutions becomes negative in the vicinity of the vacuum in the solution. The use of the conservative variables in a non characteristic method when the solution is kinetic energy rich causes the problem. This is akin to the problems with the Roe linearization studied in [231]. The interpolation of the variables creates nonphysical states in the total energy. Lowering the compression of the limiters alleviates this problem as does moving to primitive or characteristic variables for the interpolation.

C.4.4 Blast Wave Problem

The solutions are in general all quite good. The major features of this complex flow field are all depicted in the plotted density profiles (Figs. C.19–C.24). The major differences can be seen in the resolution of the contact discontinuity at $X \approx 60$, the "well" at $X \approx 75$, and the peak at $X \approx 80$.

In Fig. C.19, the CC method's major problem is the clipping of the second peak in the solution. Other features are well resolved in comparison to the other methods. The PC method (Fig. C.20) smears all the features of the flow considerably more than the CC method. The CR method is generally like the CC method with the exception of the contact discontinuity at $X \approx 60$, which is smeared much more that by the CC method. The solution is somewhat "noisier" with over/undershoots in several locations. These characteristics are duplicated in large part by the CF method (compare Figs. C.21 and C.23).

The PR and PF methods produce nearly same results. Both solutions are remarkably crisp and each feature in the flow field is sharply defined. Figures C.22 and C.24 also show the major detriment to these solution. The second peak ($X \approx 80$) significantly overshoots the "exact" solution. Nevertheless the solution found by these methods is quite good in all other respects.

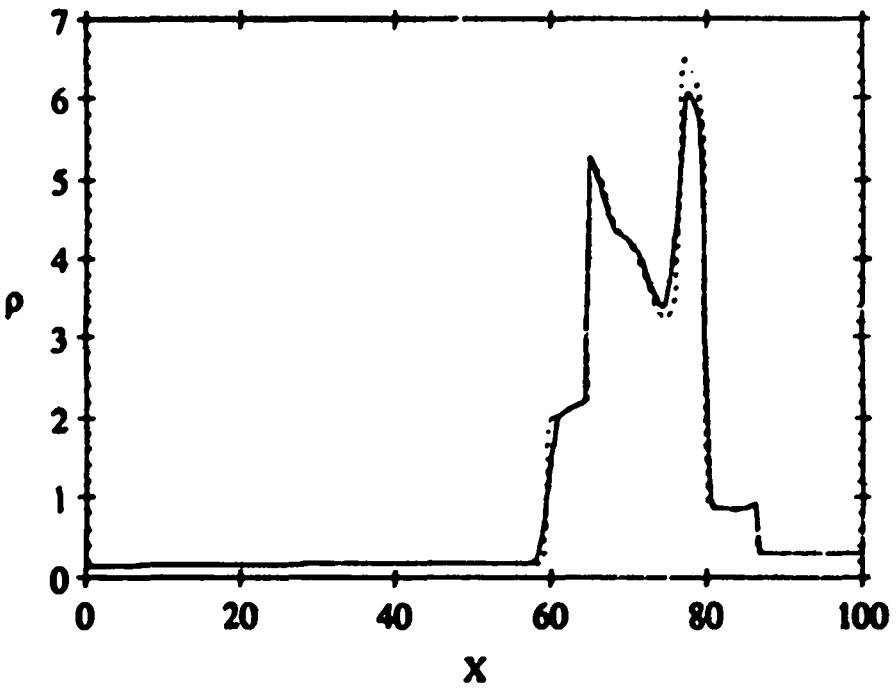


Figure C.19: The blast wave problem computed with the characteristic formulation with conservative variables. The first peak is captured very well, but the second is clipped severely. With the blast wave solution, the "exact" solution is marked by the dashed line and the approximate numerical solution by the solid line.

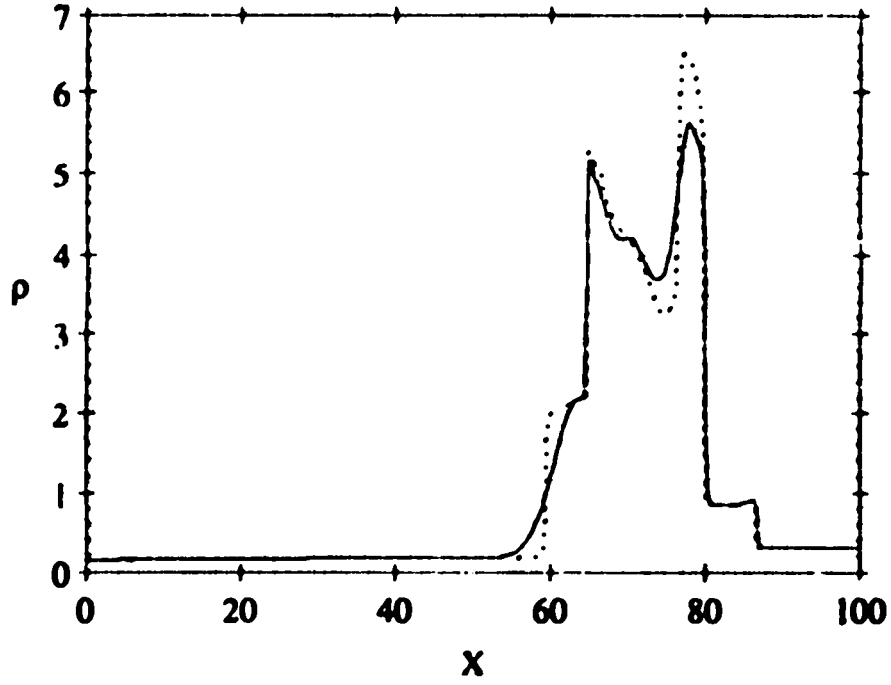


Figure C.20: The blast wave problem computed with the characteristic formulation with primitive variables. Both peaks are clipped and the contact discontinuity at $X \approx 60$ is smeared.

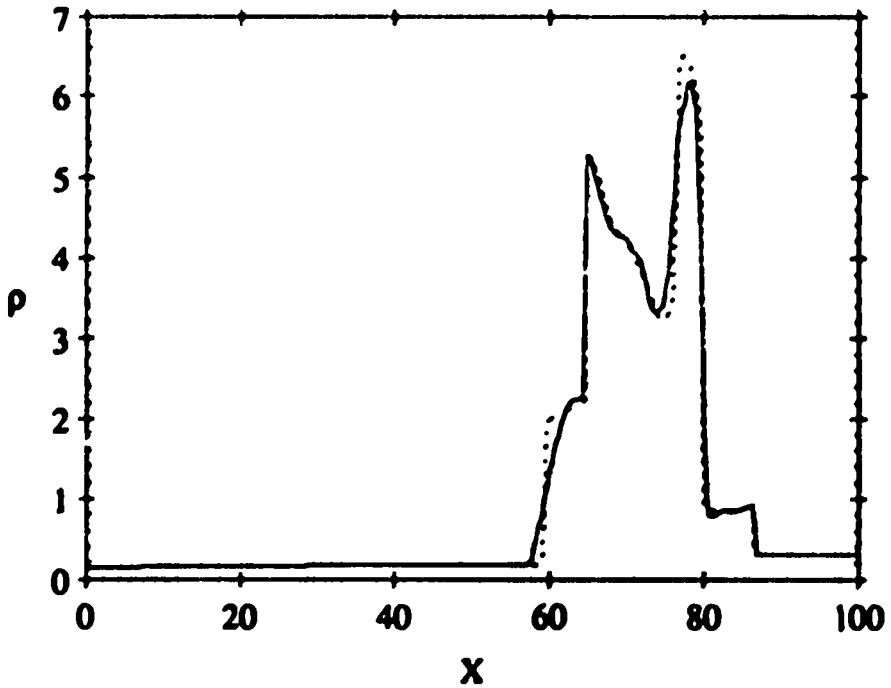


Figure C.21: The blast wave problem computed with the two-step formulation with conservative variables. This is similar to Fig. C.19, but the contact discontinuity at $X \approx 60$ is smeared significantly more.

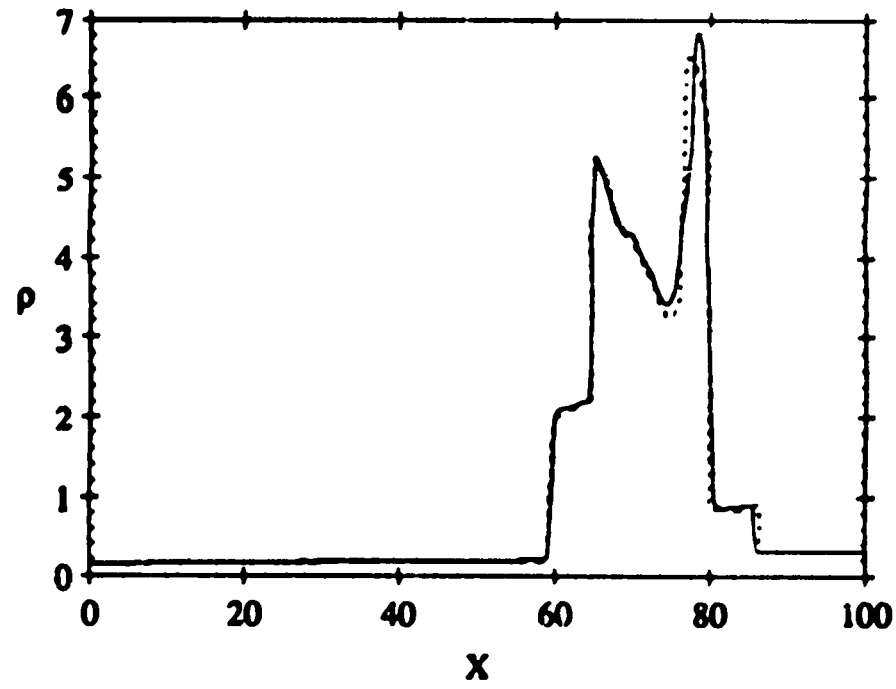


Figure C.22: The blast wave problem computed with the two-step formulation with primitive variables. This solution is highly resolved and is of high quality with the exception of the overshoot of the second peak.

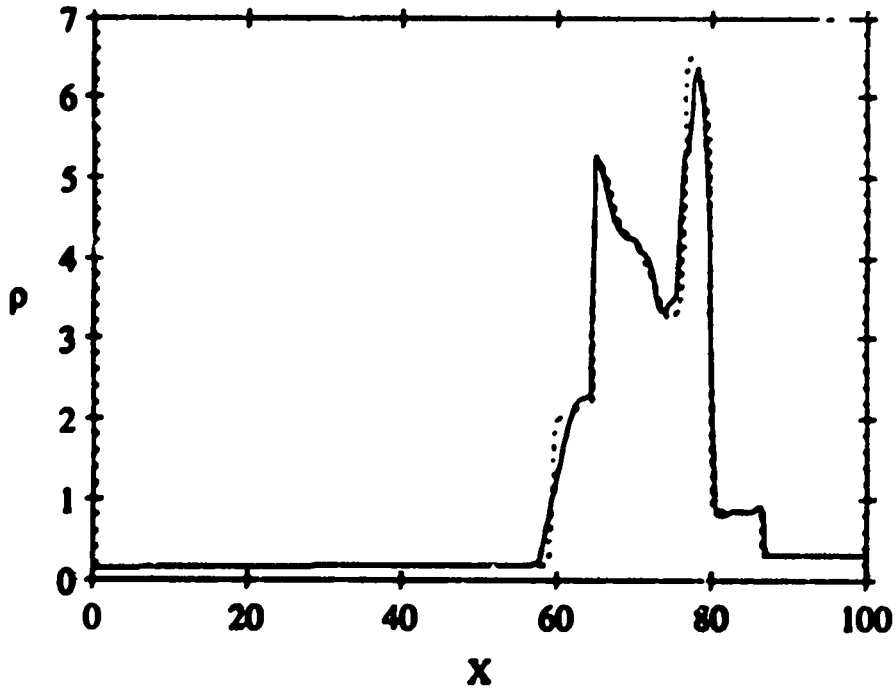


Figure C.23: The blast wave problem computed with the component-wise formulation with conservative variables. This solution is fairly well resolved, but is somewhat "noisier" than other solutions.

Table C.5: The times for the blast wave solution computation using each method

Scheme	Total Time (s)	Percentage in Reconstruction
CC	81.93	49.58
PC	79.41	49.55
CR	82.49	43.12
PR	72.01	42.57
CF	84.22	40.44
PF	69.07	40.54

C.5 Concluding Remarks

Table C.5 shows the total time taken for the blast wave solutions and the percentage of that time taken by the reconstruction of the cell-edge values[1]. In terms of economy, the PR and PF methods have clear advantages. Taking this into account with the results in mind several conclusions can be drawn. These conclusions are summarized below:

- All the methods described in the appendix produce quality results.
- When a non characteristic extension is used care must be taken in applying limiters (to not over-compress the density).
- For non characteristic extensions, the primitive variables formulation should be used.
- Non characteristic formulations using the primitive variables are lower in cost.

Another point not emphasised here has been extension to multiple dimensional problems. All of these methods can be used with a dimensional splitting method, but the two-step method has clear applicability to a purely multidimensional methods without splitting. This is clearly an advantageous feature. In sum, both of the characteristic approaches (CC and PC) are reliable and produce excellent results in all cases. The two-step primitive variable method (PR) with appropriate selection of limiters is both economical and has applicability to a multidimensional algorithm.

[1]The timings were done on a SPARCStation 2 running SunOS 4.1.1b

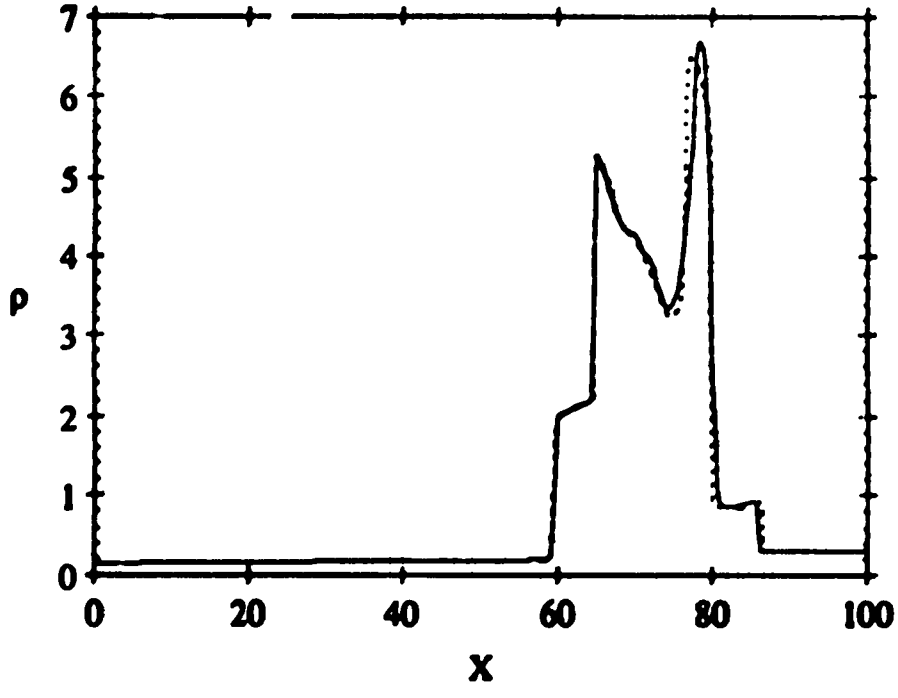


Figure C.24: The blast wave problem computed with the component-wise formulation with conservative variables. This solution is very similar to Fig. C.22.

Appendix D.

A More Robust Characteristic Reconstruction

D.1 Methodology

In [234], Colella discusses a more robust means to accomplish characteristic reconstruction. In this appendix, I show this method and explore its use.

Briefly stated, this is a modification of the methodology given earlier. For constant coefficient problems these steps lead to identical values for $\mathbf{U}_{j+\frac{1}{2},l/r}$, but as Colella comments leads to a more robust algorithm in the case of highly nonlinear problems. This method requires that we define left and right reference states, $\hat{\mathbf{U}}_{j+\frac{1}{2},l}$ and $\hat{\mathbf{U}}_{j+\frac{1}{2},r}$, respectively. These states are defined as

$$\hat{\mathbf{U}}_{j+\frac{1}{2},l} = \mathbf{U}_j + \frac{1}{2}\left(1 - \max\left(\lambda_j^K, 0\right)\right)\widetilde{\Delta_j \mathbf{U}}\,, \tag{D.1a}$$

and

$$\hat{\mathbf{U}}_{j+\frac{1}{2},r} = \mathbf{U}_j - \frac{1}{2}\left(1 - \min\left(\lambda_{j+1}^1, 0\right)\right)\widetilde{\Delta_{j+1}\mathbf{U}}\,. \tag{D.1b}$$

Here, the eigenvalues, λ^k, have been arranged in increasing order from $\lambda^1 \cdots \lambda^K$. These reference states are then used in defining the cell-edge values as

$$\mathbf{U}_{j+\frac{1}{2},l} = \hat{\mathbf{U}}_{j+\frac{1}{2},l} + \frac{1}{2}\sum_{k:\lambda^k>0} r^k\left(\lambda_j^K - \lambda_j^k\right)\widetilde{\Delta_j \alpha^k}\,, \tag{D.2a}$$

and

$$\mathbf{U}_{j+\frac{1}{2},r} = \hat{\mathbf{U}}_{j+\frac{1}{2},r} + \frac{1}{2}\sum_{k:\lambda^k<0} r^k\left(\lambda_{j+1}^1 - \lambda_{j+1}^k\right)\widetilde{\Delta_{j+1}\alpha^k}\,. \tag{D.2b}$$

All the above terms where defined in Chapter C. One would expect this method to be slightly more diffusive than the usual reconstruction because of the lack of extrapolation of the linear profile for eigenvalues that do not propagate toward the cell edge.

D.2 Results

I compare the above described method with the more straight forward algorithm used throughout this research. To do this I use the same four test problems described in Chapter A. To simplify comparison on the density and velocity profiles are studied.

For Sod's problem, the more robust algorithms sole improvement seems to be in the velocity profile where the "bump" experienced with the usual algorithm near the end of the rarefaction wave has disappeared. This is shown in Fig. D.1. The L_1 error for density is also slightly better.

With Lax's problem, the difference is barely perceptible. Figure D.2 shows that the two solutions are nearly identical. The L_1 error norm for density is slightly worse for the robust reconstruction.

Again for the vacuum problem as with Lax's problem, the two solutions are not greatly different, although the robust reconstruction appears to be more diffusive. As Fig. D.3 shows, near the vacuum in the solution, the robust reconstruction shows more artificial diffusion.

Figure D.4 shows the solutions for the two methods on the blast wave problem. The solutions were computed with 500 grid points. Only the region of wave interactions is shown. Again, as shown in this figure, the solutions are very similar.

While the robust reconstruction does not have any detrimental effects on the solution (save a little artificial diffusion), except in the case of Sod's problem, it does not improve the solution. It is also somewhat more expensive than the usual reconstruction, although this cost is not particularly high.

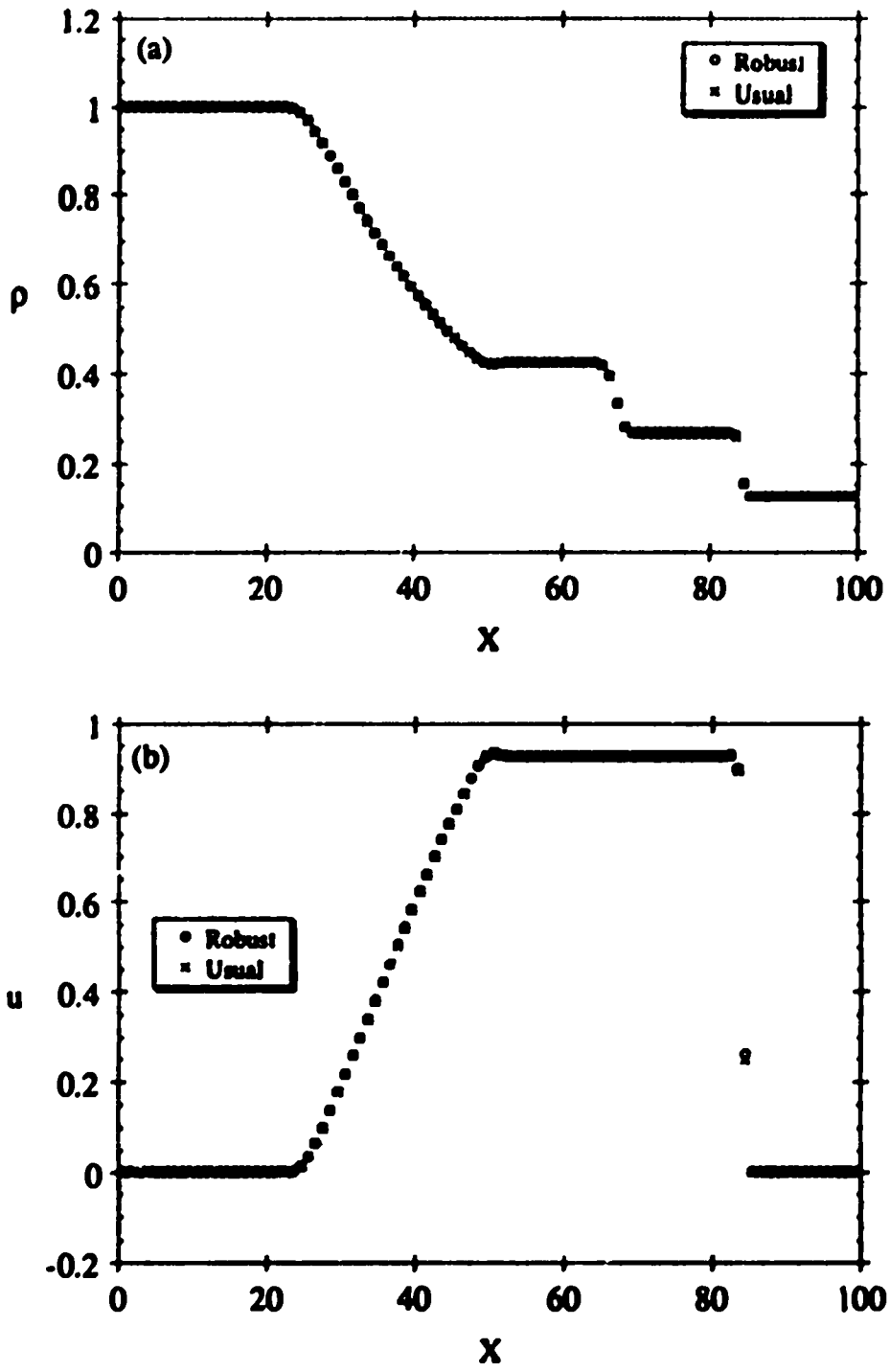


Figure D 1: The density and velocity solutions to Sod's problem using both the usual and robust reconstruction methods.

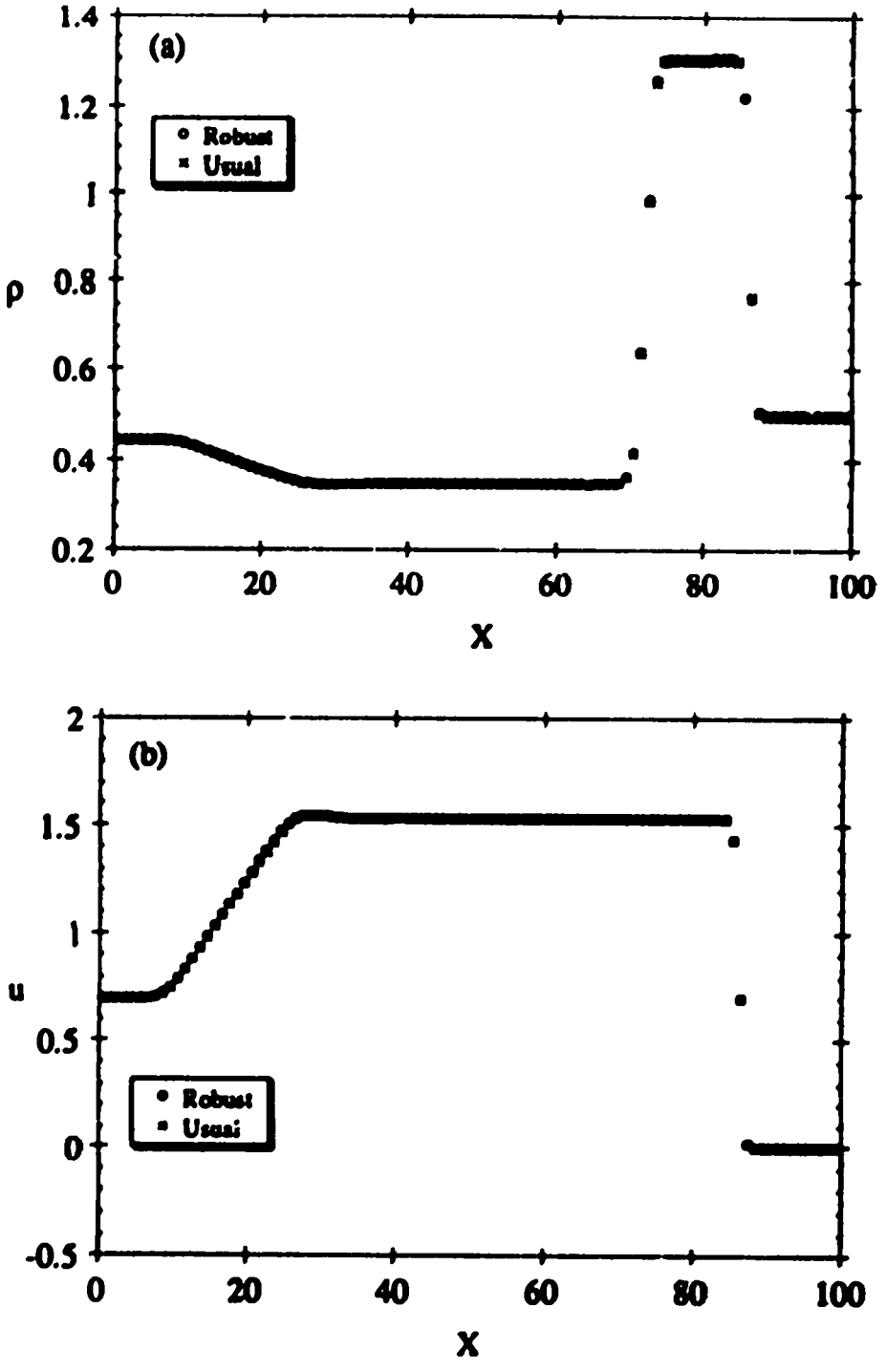


Figure D.2: The density and velocity solutions to the vacuum problem using both the usual and robust reconstruction methods.

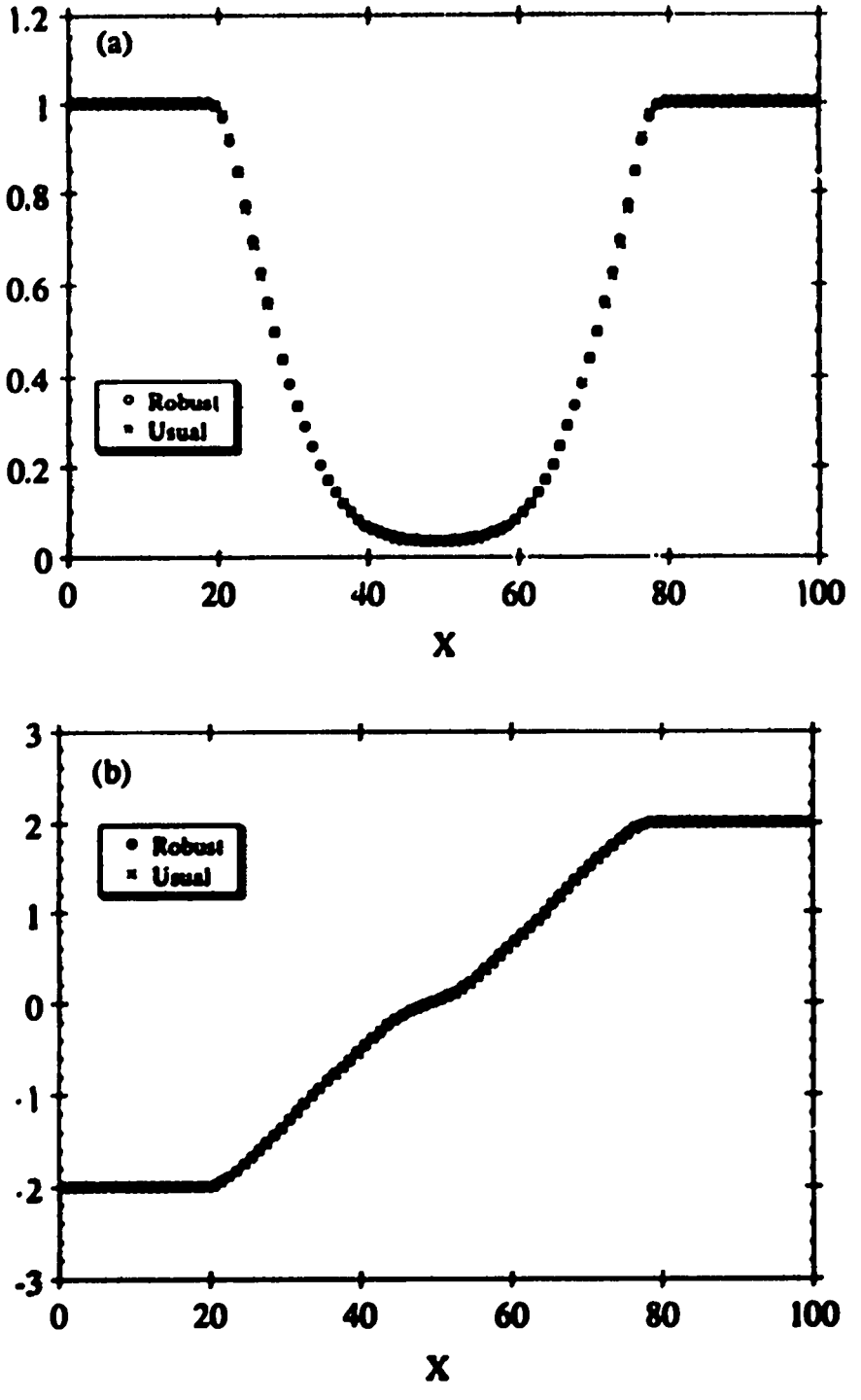


Figure D.3: The density and velocity solutions to the vacuum problem using both the usual and robust reconstruction methods.

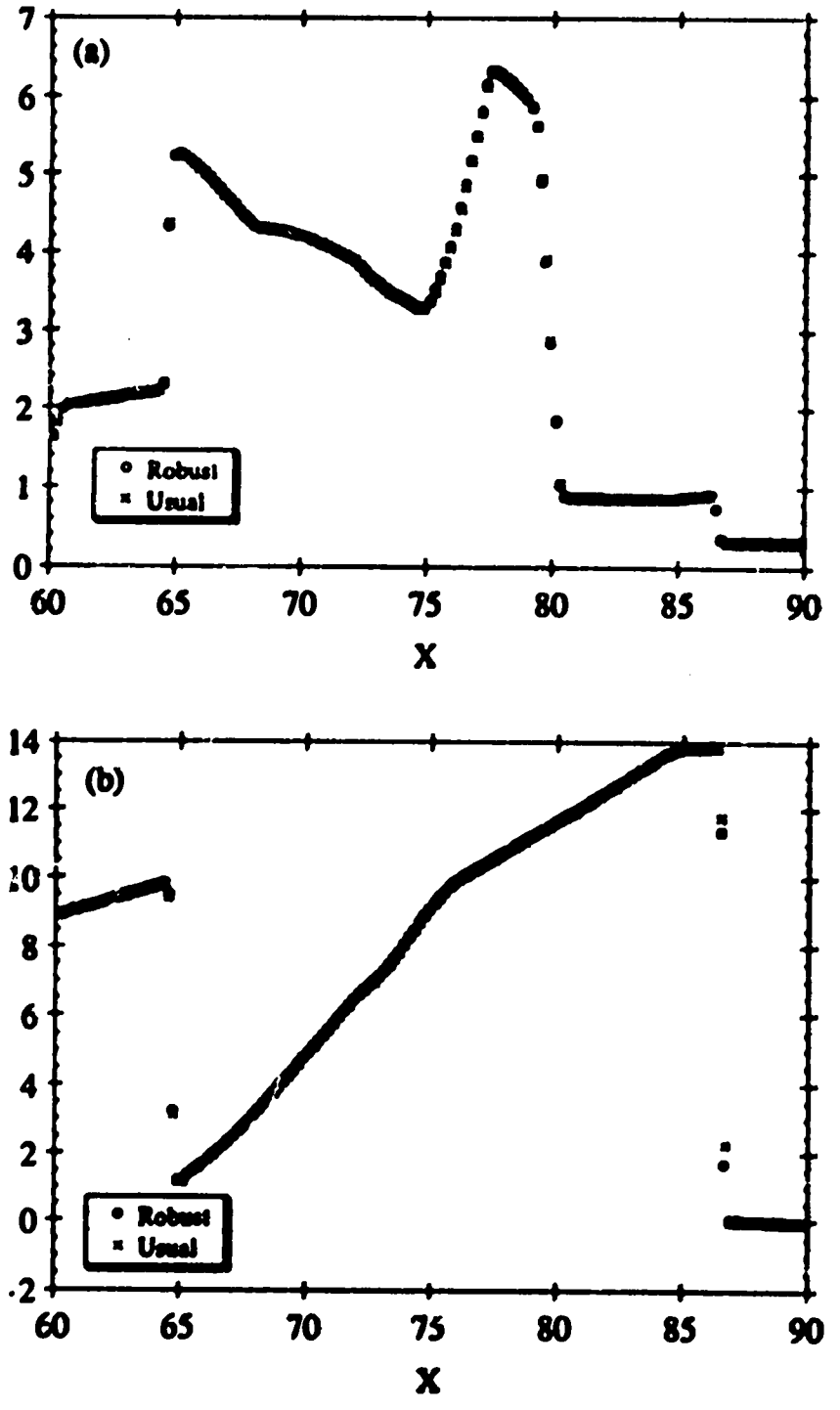


Figure D.4: The density and velocity solutions to the blast wave problem using both the usual and robust reconstruction methods.

Appendix E.
Neo-Classical Upwind Type Methods

Here I briefly explore the types of solutions that arise from the solution of modified flux and symmetric TVD schemes without limiters. The schemes can be derived from those schemes by considering what the fluxes would be for the various sample gradients used in the limiters. This gives three separate schemes for the modified flux type of method: upwind, antiupwind and centered (or average of the other two). For the symmetric method, four schemes arise: upwind (Beam Warming), centered (Lax-Wendroff), antiupwind and average.

The results for these methods on the scalar advection of a square wave for 100 time steps at $\nu = 0.5$ can be seen in Figs. E.1 and E.2. Each of the solutions is second-order accurate and shows distinct dispersive effects. For the modified flux type of scheme, the upwind and antiupwind errors are opposite in orientation and the centered solution is superior. For the symmetric scheme, surprisingly, the antiupwind method followed by the average method seem to be superior in terms of oscillation control.

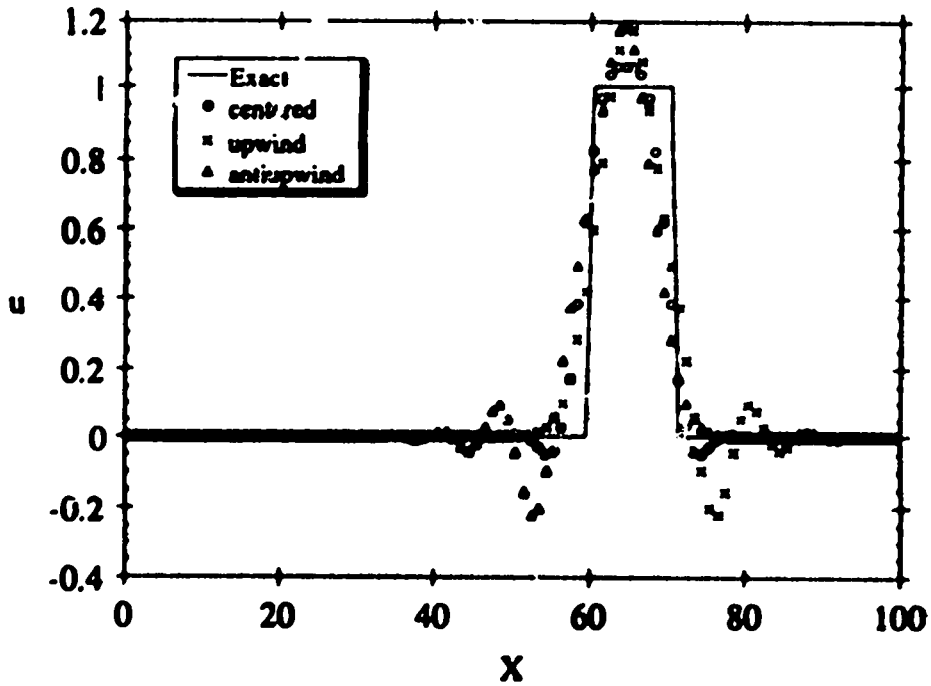


Figure E.1: The solutions for the neo-classical modified flux upwind schemes on the scalar advection of a square wave ($a = 1$ and $\nu = 0.5$).

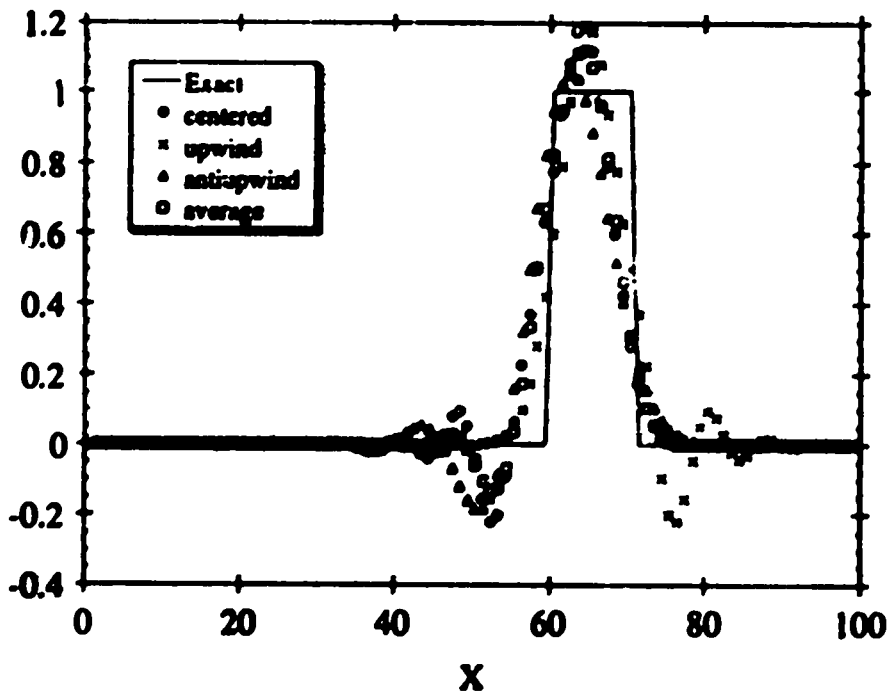


Figure E.2: The solutions for the neo-classical symmetric upwind schemes on the scalar advection of a square wave ($a = 1$ and $\nu = 0.5$).

Appendix F.

Extension of High Resolution Schemes to Multiple Dimensions

F.1 Introduction

Methods for numerically integrating conservation laws are best understood in one dimension. Because of this, schemes are most often developed and thoroughly tested in one dimension. High-resolution schemes are no exception to this rule. In some cases, a good one-dimensional method cannot be generalized to multiple dimensions because of assumptions made in their derivation. Fortunately, this is not always true, although the one-dimensional methods are always somewhat limited when used in multiple-space dimensions.

The more straightforward methods for the multidimensional advection algorithms are developed in physically or logically rectangular coordinates. Finite element methods and more general finite volume methods [35, 36] can be defined for more general geometries. The problem with these methods is that the theoretical support in multidimensions is somewhat lacking. A perfect example of this difficulty is with Riemann solvers. Multidimensional Riemann solvers are an active topic of research [228, 235, 236, 237, 229], but in one dimension, Riemann solvers are well developed. Typically, Riemann solvers are used in an operator splitting fashion [156] where at each cell interface the multidimensional problem is reduced to a one-dimensional problem. These are then pieced together to give a multidimensional algorithm [234]. As is discussed shortly, the advent of multidimensional Riemann solvers do not cure all the problems associated with the solution of multidimensional problems with high-resolution upwind methods.

A common approach to achieving high-resolution methods is the use of flux or slope limiters. For one space dimension, limiters are well developed, but for more than one dimension, their development is somewhat less. One aspect to multidimensional limiters is that they require the use of more ample gradients than their one-dimensional counterparts. As discussed in Chapter 8, the more arguments given to a limiter, the lower its resolution simply because of the minimum principle used. Multidimensional limiters have been given by [238, 139, 239].

In this appendix, I attempt to see what some of these limitations are and what methodology is best suited to the task. The appendix is organized into nine sections: an introduction, a description of the first-order methods, the test problems, and the first order results. This is followed by a description of the basic high-resolution

method and its extension to multiple dimensions. After this, the results of the high-resolution methods in two space dimensions is given. Following that discussion is a brief description of the impact of limiter selection on the results. Finally, some closing remarks are made.

F.2 First-Order Methods in Multiple Spatial Dimensions

In this appendix I am interested in solving the following equation,

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} + \frac{\partial g(u)}{\partial y} = 0 , \tag{F.1}$$

where $f(u) = au$ and $g(u) = bu$. A conservative differencing of this equation is

$$u_{i,j}^{n+1} = u_{i,j}^{n} - \sigma_x \left(f_{i+\frac{1}{2},j,lr} - f_{i-\frac{1}{2},j,lr} \right) - \sigma_y \left(g_{i,j+\frac{1}{2},bl} - g_{i,j-\frac{1}{2},bl} \right) , \tag{F.2}$$

where $\sigma_x = \Delta t / \Delta x$ and $\sigma_y = \Delta t / \Delta y$.

In each of the methods discussed in this appendix, the cell-edge flux at cell edge $i + \frac{1}{2}, j$ are defined by the following approximate Riemann solver for scalar wave equations

$$f_{i+\frac{1}{2},j,lr} = \frac{1}{2} \left[a \left(u_{i+\frac{1}{2},j,l} + u_{i+\frac{1}{2},j,r} \right) - |a| \left(u_{i+\frac{1}{2},j,r} - u_{i+\frac{1}{2},j,l} \right) \right] , \tag{F.3}$$

where a is the velocity in the x-direction at the cell edge and the subscript l refers to the value to the left of the cell edge, r to the right and lr is the interface value. Similarly, the flux is the y-direction at cell edge $i, j + \frac{1}{2}$ is

$$g_{i,j+\frac{1}{2},bl} = \frac{1}{2} \left[b \left(u_{i,j+\frac{1}{2},b} + u_{i,j+\frac{1}{2},t} \right) - |b| \left(u_{i,j+\frac{1}{2},t} - u_{i,j+\frac{1}{2},b} \right) \right] , \tag{F.4}$$

where b is the velocity in the y-direction at the cell edge and the subscript b refers to the value at the bottom of the cell edge, t to the top, and bt is the interface value. By defining the cell-edge values, I then define the scheme.

For the first-order schemes, the value at the cell edges are given by the value of the variable in each cell for instance

$$u_{i+\frac{1}{2},j,l} = u_{i,j} , \tag{F.5a}$$

and

$$u_{i+\frac{1}{2},j,r} = u_{i+1,j} , \tag{F.5b}$$

and other cell-edge values defined in a similar fashion. The simplest scheme is then

from the conservation form, (F.2).

Another common form uses dimensional splitting [156] usually implemented with Strang splitting [240, 241]. This method pieces together one-dimensional solutions into a multidimensional solution. For two dimensions, I can order the solution in two ways as either

$$u_{i,j}^{n+1} = \mathcal{L}_x \mathcal{L}_y \left(u_{i,j}^n \right) , \tag{F.6a}$$

or

$$u_{i,j}^{n+1} = \mathcal{L}_y \mathcal{L}_x \left(u_{i,j}^n \right) . \tag{F.6b}$$

Here the operator $\mathcal{L}_x \mathcal{L}_y \left(u_{i,j} \right)$ would be carried out in two steps, the first being

$$u_{i,j}^* = \mathcal{L}_y \left(u_{i,j}^n \right) . \tag{F.7a}$$

and the second being

$$u_{i,j}^{n+1} = \mathcal{L}_x \left(u_{i,j}^* \right) . \tag{F.7b}$$

with $\mathcal{L}_y \mathcal{L}_x \left(u_{i,j}^n \right)$ defined in a similar manner. The function $\mathcal{L}_x \left(u_{i,j}^\theta \right)$ is defined as

$$\mathcal{L}_x \left(u_{i,j}^\theta \right) = u_{i,j}^\theta - \sigma_x \left(f_{i+\frac{1}{2},j,k}^\theta - f_{i-\frac{1}{2},j,k}^\theta \right) , \tag{F.8a}$$

and $\mathcal{L}_y \left(u_{i,j}^\theta \right)$ is defined as

$$\mathcal{L}_y \left(u_{i,j}^\theta \right) = u_{i,j}^\theta - \sigma_y \left(g_{i,j+\frac{1}{2},k}^\theta - g_{i,j-\frac{1}{2},k}^\theta \right) . \tag{F.8b}$$

Strang [240] showed that if the order of evaluation is alternated, errors cancel to second-order in time (also see LeVeque [40, Chapter 18]) thus the implemented order of evaluation for two time steps is

$$u_{i,j}^{n+2} = \mathcal{L}_y \mathcal{L}_x \mathcal{L}_x \mathcal{L}_y \left(u_{i,j}^n \right) . \tag{F.9}$$

The use of this with Godunov's method defines the split Godunov method.[1]

Colella defines a third choice for multidimensional extensions of one-dimensional methods. He calls these corner transported upwind (CTU) methods, a term I use here. The basic geometric idea is shown in Fig. F.1. This is a two-step method that defines time-centered values for the cell edges and uses these to compute the advance time cell-centered values. The first step of the method computes a time- centered value for a cell edge based on the characteristics traced from the corners of that cell

[1] The use of Strang splitting is not necessary with first-order methods, but is really needed for second-order methods

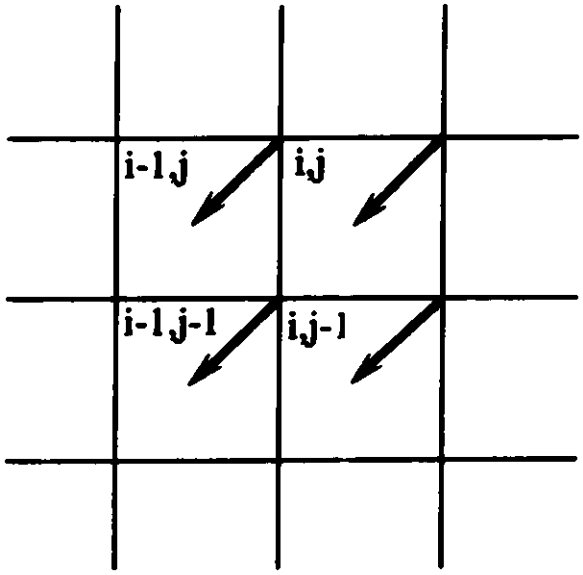


Figure F.1: A diagram showing the trace of characteristics back from the cell corner of cell (i, j) with both velocities being positive

edge. For the x-direction cell edge, this gives

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j;l} = u^n_{i,j} - \frac{\sigma_y}{2}\left(g^n_{i,j+\frac{1}{2};bt} - g^n_{i,j-\frac{1}{2};bt}\right) , \tag{F.10a}$$

and for the y-direction cell edge

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j;l} = u^n_{i,j} - \frac{\sigma_x}{2}\left(f^n_{i+\frac{1}{2},j;lr} - f^n_{i-\frac{1}{2},j;lr}\right) . \tag{F.10b}$$

The fluxes are computed by some means, in this case a Godunov flux as described above. The final time-advanced solution is computed from

$$u^{n+1}_{i,j} = u^n_{i,j} - \sigma_x\left(f^{n+\frac{1}{2}}_{i+\frac{1}{2},j;lr} - f^{n+\frac{1}{2}}_{i-\frac{1}{2},j;lr}\right) - \sigma_y\left(g^{n+\frac{1}{2}}_{i,j+\frac{1}{2};bt} - g^{n+\frac{1}{2}}_{i,j-\frac{1}{2};bt}\right) , \tag{F.10c}$$

which uses the CTU-time-centered values to define the Godunov fluxes.

Before continuing, some comments concerning stability should be made. Classical stability analysis applies to the above schemes. For the split and CTU Godunov schemes the stability limit is

$$\max_{i,j}\left(\nu_x, \nu_y\right) \leq 1 , \tag{F.11a}$$

and for the unsplit Godunov scheme

$$\nu_x + \nu_y \leq \quad , \tag{F.11b}$$

where $\nu_x = |a|\,\sigma_x$ and $\nu_y = |b|\,\sigma_y$.

F.3 Test Cases and Problem Setup

In this appendix, I consider three test problems as initial conditions to the multidimensional scalar wave equation. The equation I solve is

$$\frac{\partial u}{\partial t} + \frac{\partial (a(y)\, u)}{\partial x} + \frac{\partial (b(x)\, u)}{\partial y} = 0 \tag{F.12a}$$

where

$$a(y) = -\omega (y - y_c) \ , \tag{F.12b}$$

and

$$b(x) = \omega (x - x_c) \ , \tag{F.12c}$$

with $\omega = 0.1$, $x_c = 50$ and $y_c = 50$. At $t = 20\pi$ the field has rotated once. The overall domain is $[x_0, x_n] \times [y_0, y_n] = [0, 100] \times [0, 100]$. This problem setup follows Zalesak [62] and Munz [181]. I use a time step size of $20\pi/628$ so that the profile revolves once in 628 time steps.

The first problem is defined by Smolarkiewicz [242] as the cone problem. The initial conditions and exact solution are shown in Fig. F.2. The cone is centered at (50, 75) with a height of unity and a radius of 15. This problem should show how the solutions maintain local extrema and shape during advection. For the cone and the slotted cylinder problems, the figures are only shown a 50×50 portion of the grid in order to concentrate on the solution.

The second problem is the slotted cylinder problem introduced by Zalesak in [62]. This problem has been used by a number of researchers [181, 93, 242] to test multidimensional advection schemes. The cylinder is centered at (50, 75) and has a height of unity and a radius of 15. A slot is cut out of the cylinder at its lower center leaving a "bridge" with a maximum width of 5. This problem highlights the performance of the methods on contact discontinuities showing their numerical diffusion. Figure F.3 shows the initial condition for the slotted cylinder.

F.4 First Order-Results

In this section I discuss the results of using the first-order methods on the rotating cone and slotted cylinder problems after one rotation. In general, the solutions all have similar properties and results. Graphically speaking, the solutions are nearly identical. This is shown by looking at Figs. F.4, F.6, and F.8 for the cone problem and Figs. F.5, F.7, and F.9 for the slotted cylinder. All these solutions show exceedingly poor resolution of the solution and the original profile is nearly indistinguishable.

The results for all the methods discussed in this appendix are given in several tables. Table F.1 shows the computer time used in producing each solution. It is notable that the CTU-Godunov method uses half again as much time as the split

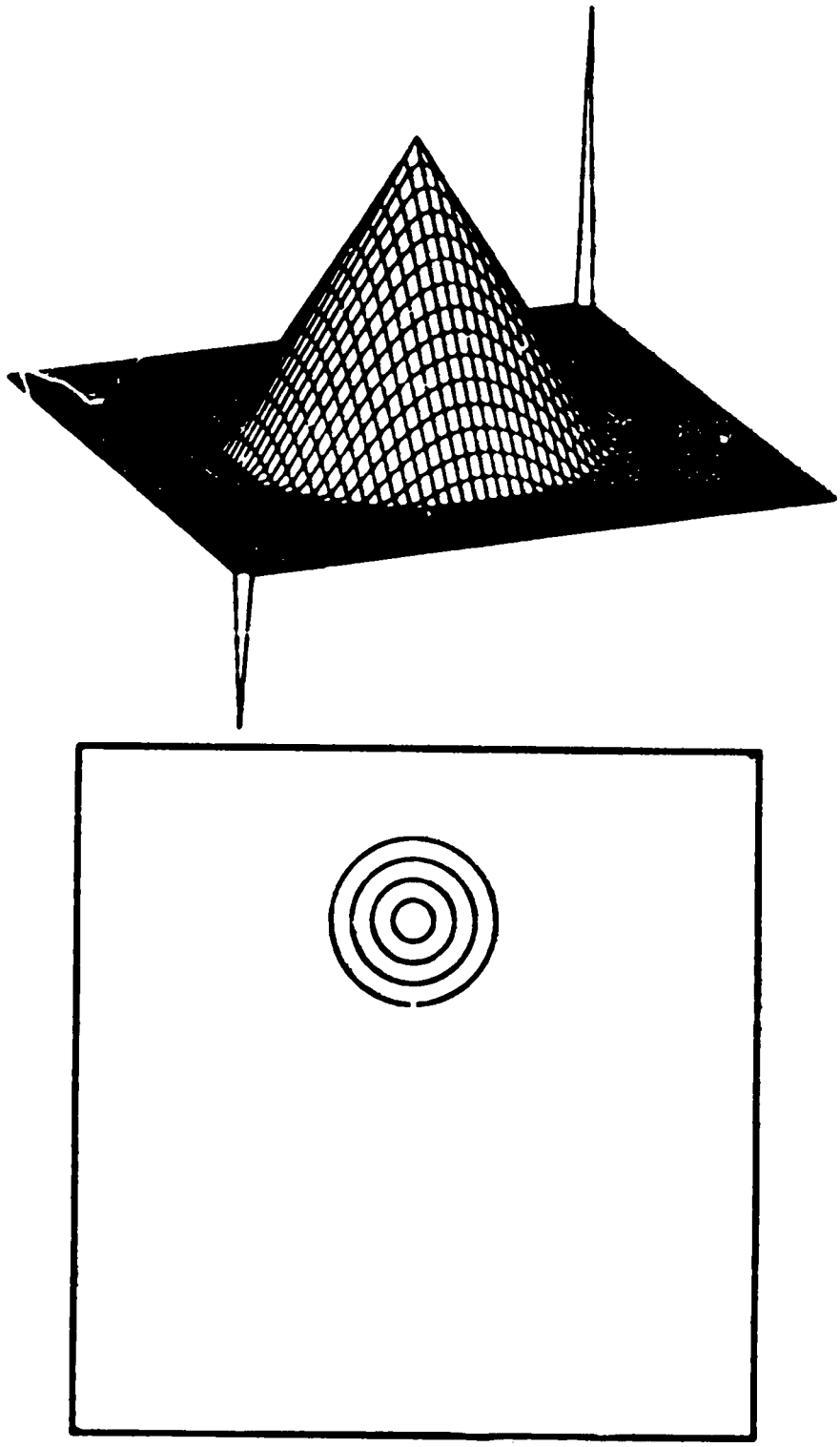


Figure F.2: Initial condition and exact solution after n rotations for the cone problem. The spike in the upper right hand corner of the upper figure is set equal to 1 and the spike in the lower left hand corner equal $-\frac{1}{2}$.

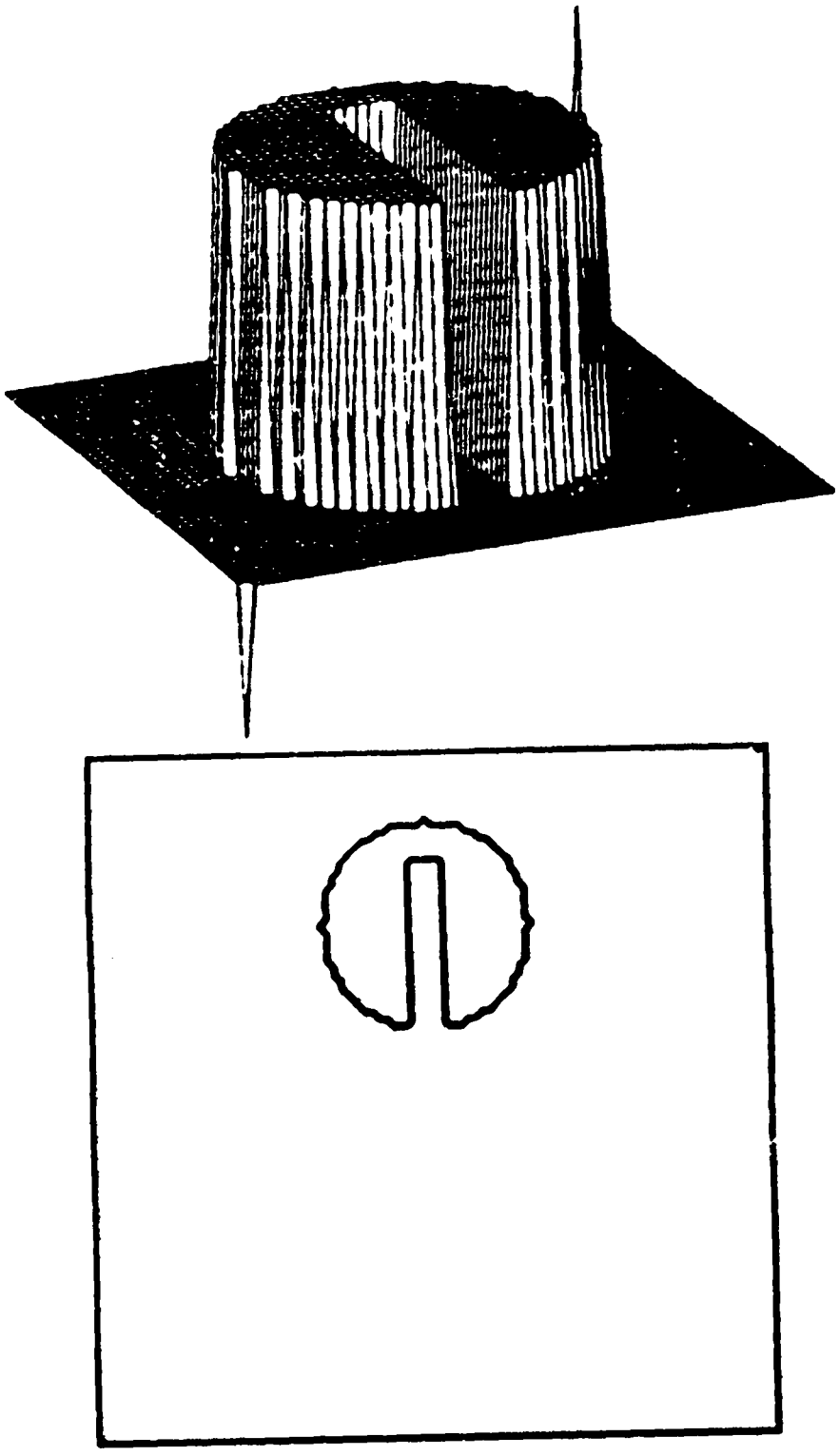


Figure F.3: Initial condition and exact solution after n rotations for the slotted cylinder problem.

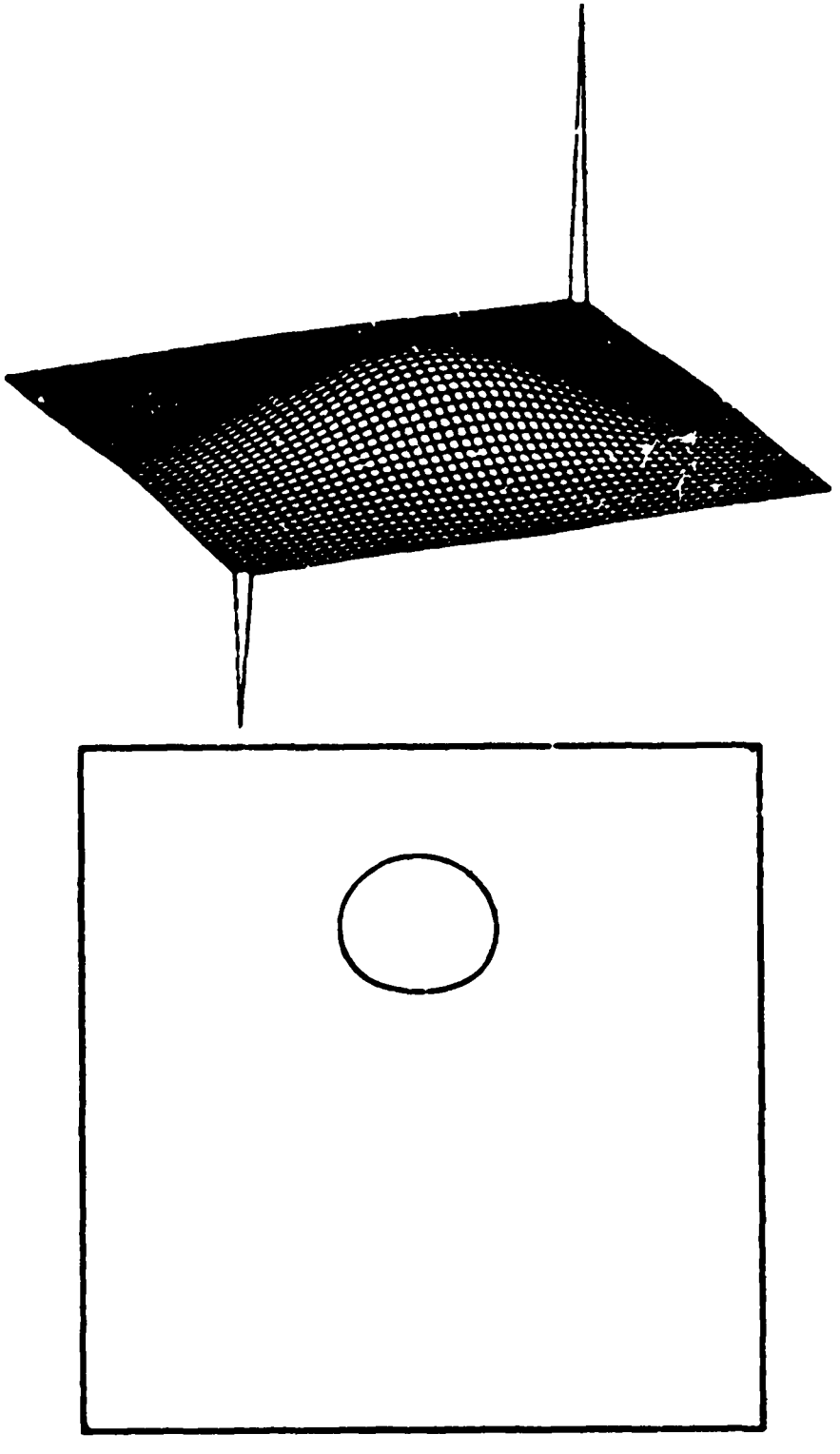


Figure F.4: The split Godunov method solution for the rotating cone shows the excessive diffusion of this method.

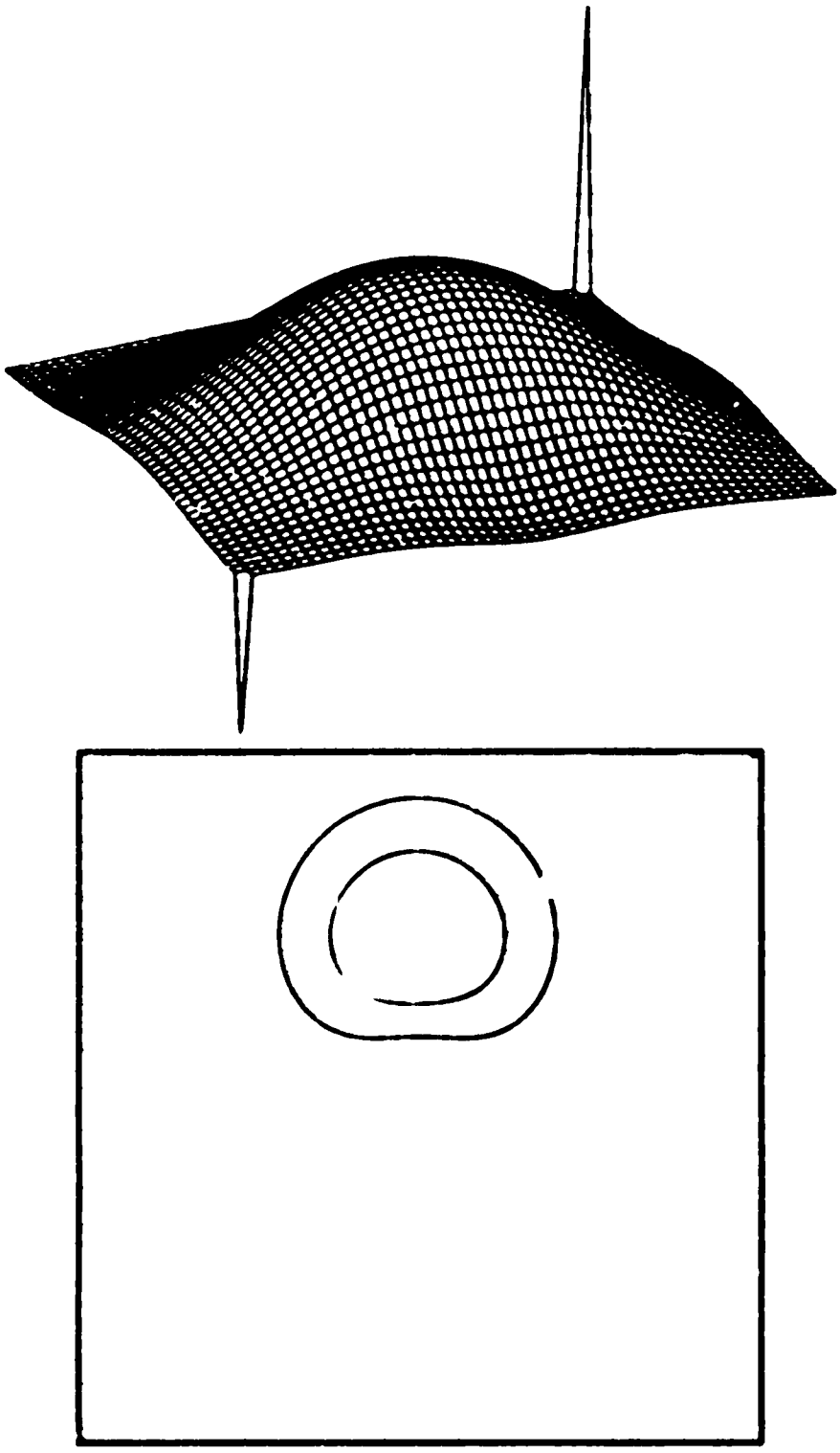


Figure F.5: The split Godunov method solution for the rotating slotted cylinder shows the excessive diffusion of this method.

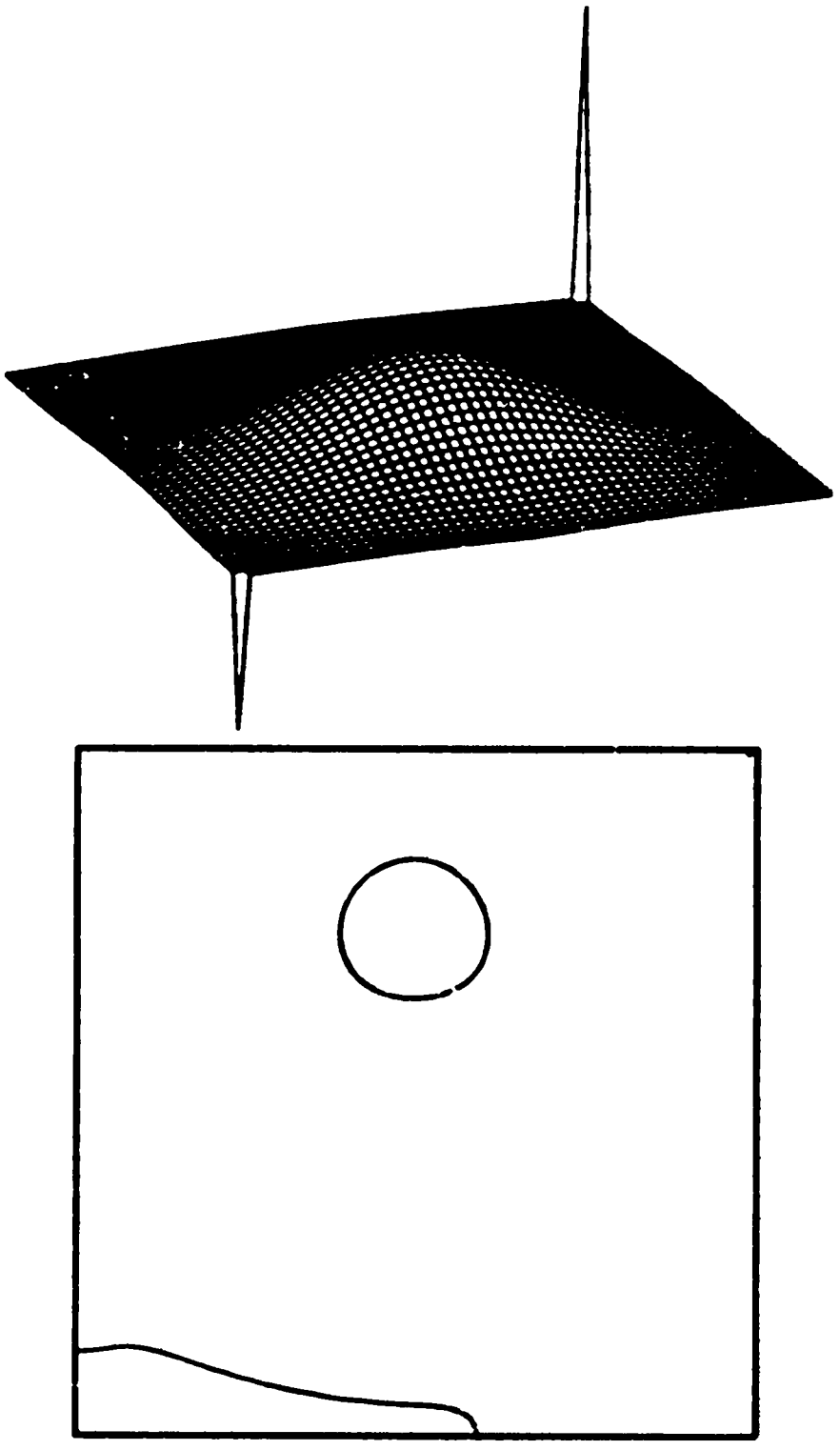


Figure F.6: The unsplit Godunov method solution for the rotating cone shows the excessive diffusion of this method.

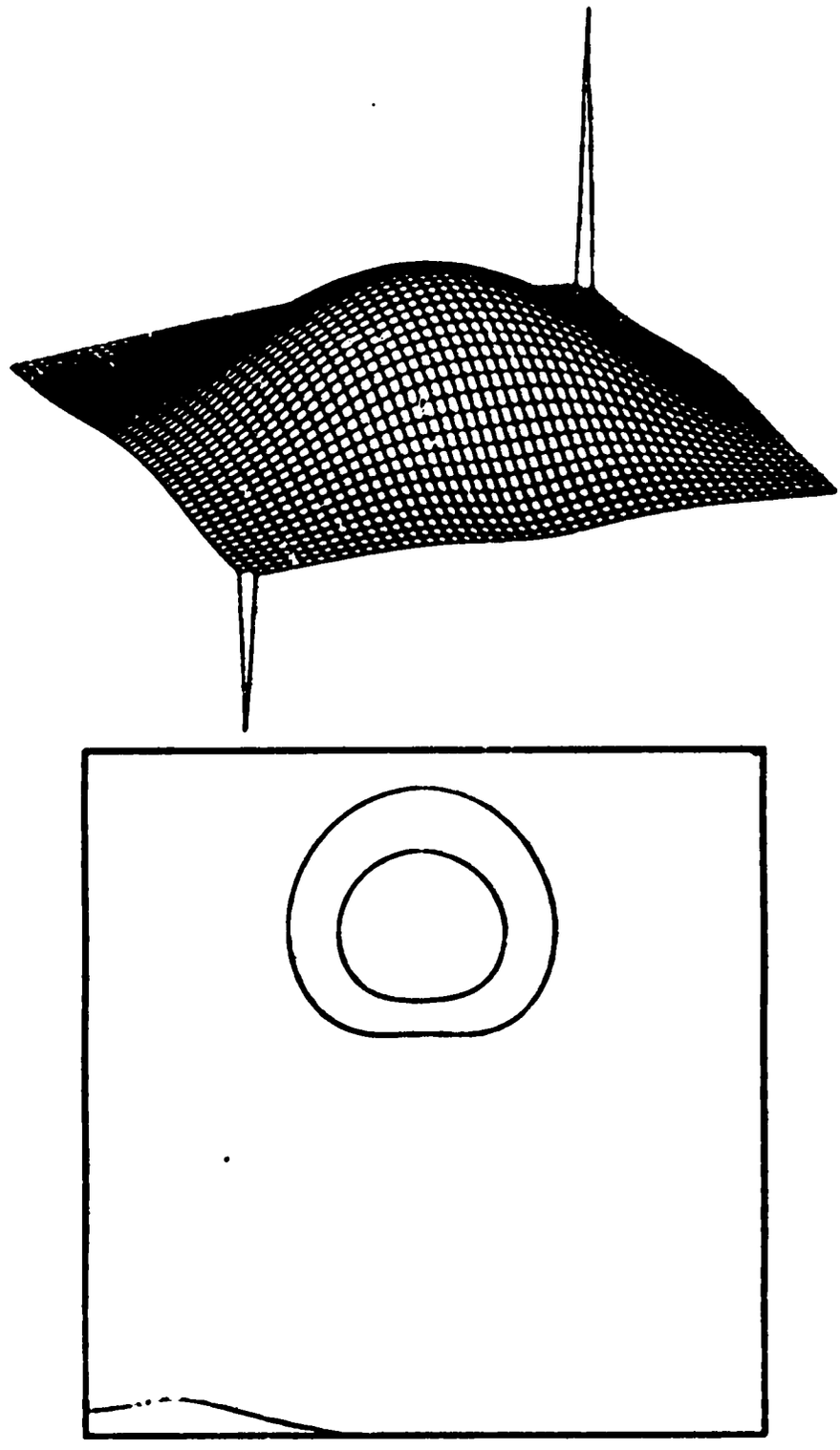


Figure F.7: The unsplit Godunov method solution for the rotating slotted cylinder shows the excessive diffusion of this method.

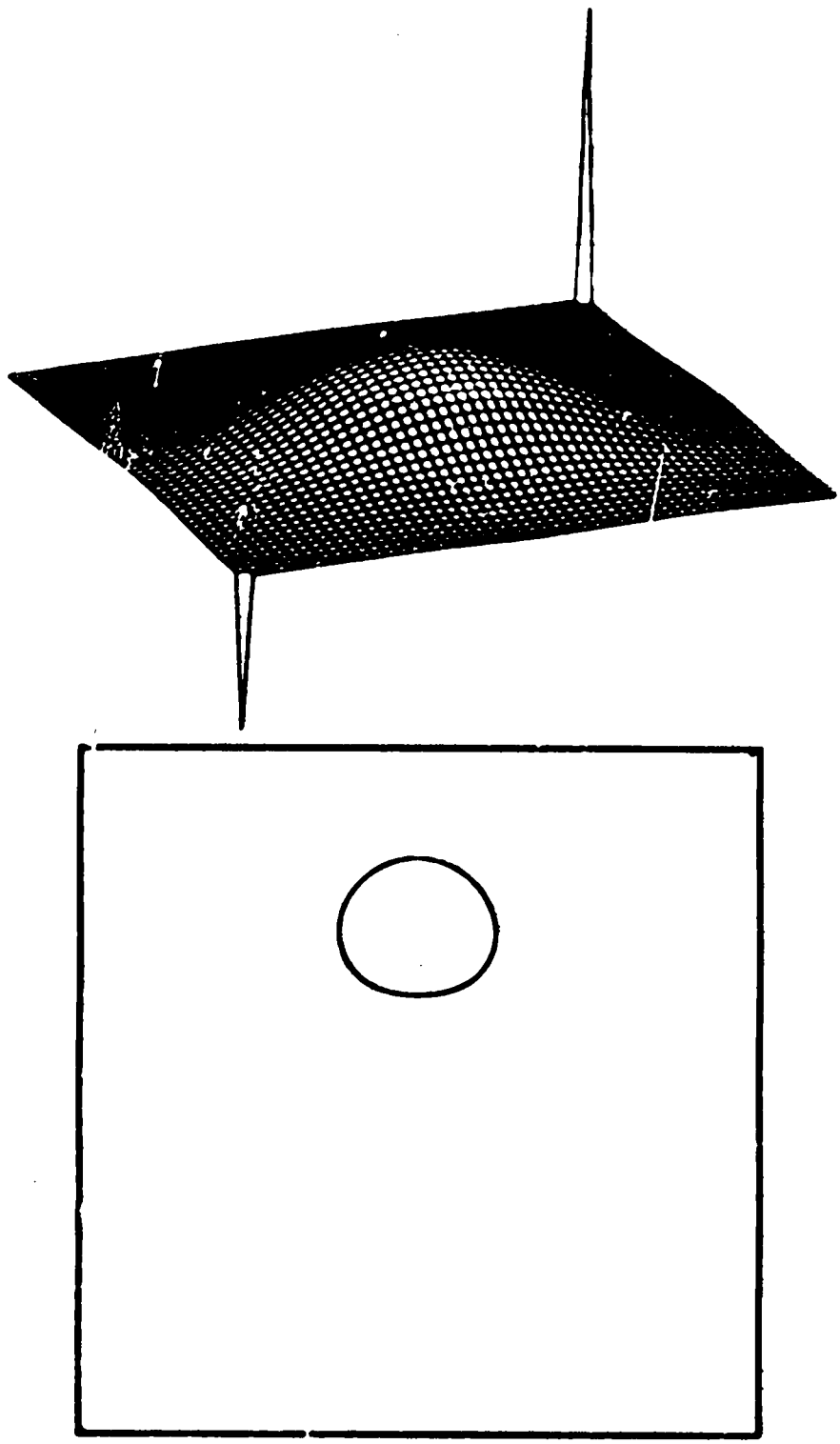


Figure F.8: The CTU-Godunov method solution for the rotating cone shows the excessive diffusion of this method.

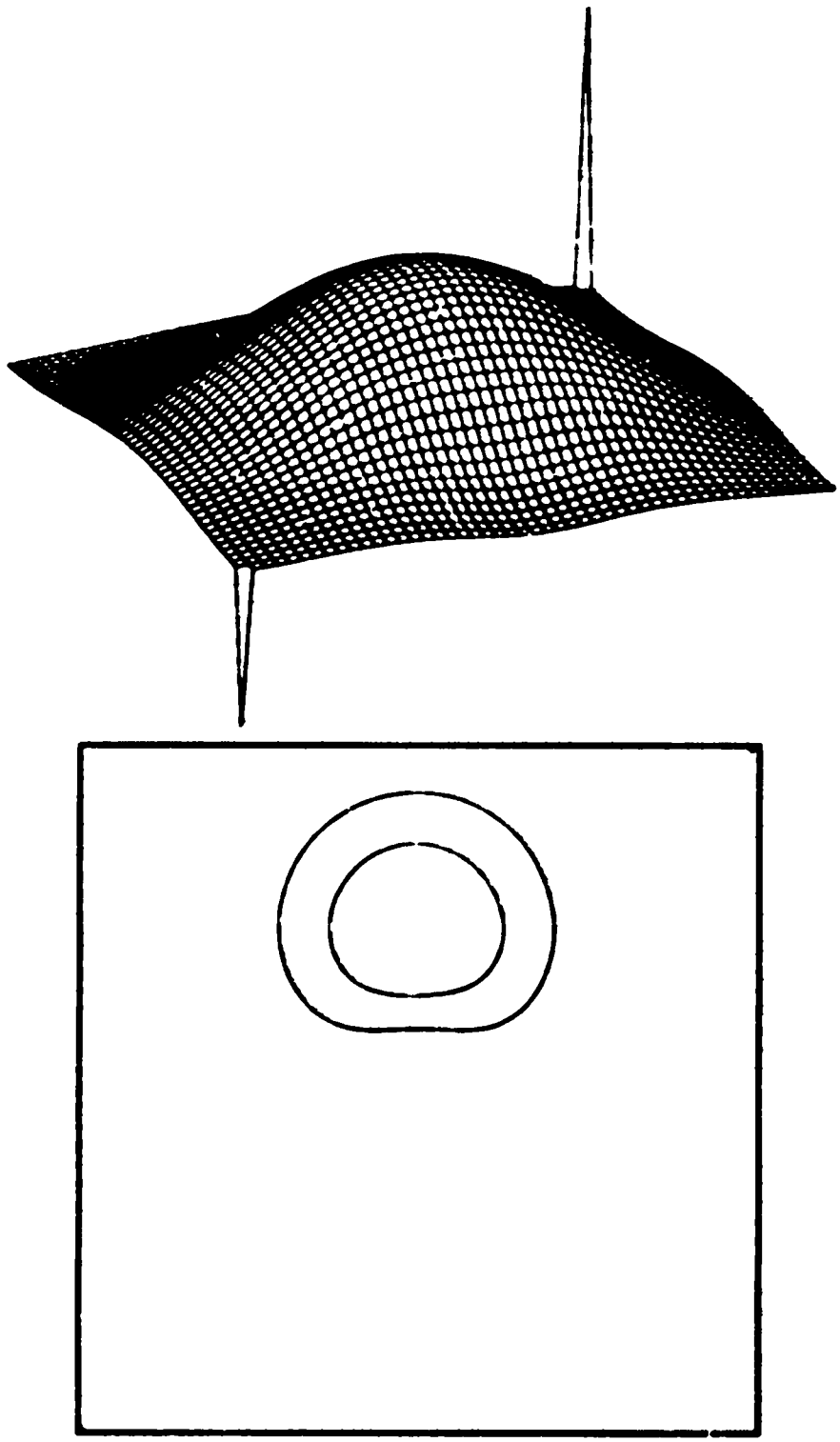


Figure F.9. The CTU-Godunov method solution for the rotating slotted cylinder shows the excessive diffusion of this method.

Table F.1: Computer time used for the solution of a problem using each method through six rotations (CFT 1.14 on a Cray X-MP4/16 with a CTSS operating system).

Scheme	CPU Time (s)	Total Time (s)
Split Godunov	27.975	41.372
Unsplit Godunov	27.640	40.905
CTU Godunov	42.455	60.256
Lax-Wendroff	39.043	55.545
Split HOG	49.913	71.684
Unsplit HOG	48.943	70.346
CTU HOG/Godunov	73.487	134.891
CTU HOG	63.542	124.737
Runge-Kutta HOG	70.885	101.848
Hancock-van Leer HOG	58.656	117.215

Godunov method to achieve nearly the same result. The times for the split and unsplit Godunov solutions are nearly equal. Table F.2 gives the solution minimums and maximums for all methods after one rotation of the cone. The split Godunov solution is slightly better than the other solutions, and all three methods are monotonic. Table F.3 shows that the slotted cylinder results yield similar conclusions.

F.5 High-Resolution Methods

This section explores methods used to improve the above results while staying within the basis of one-dimensional methods as a basic building block. Below I show the basic scheme used in the study and introduce the methods of extension to multiple dimensions.

F.5.1 The Basic One-Dimensional High-Resolution Method

To set the high-order Godunov (HOG) methods tested in this appendix on equal footing, all methods use the same basic one-dimensional method as a basis. This method is a simple second-order method defined by the following piecewise polynomial function in the x-direction

$$P_{i,j}(x) = u_{i,j} + \widetilde{\Delta_x u}\frac{x - x_{i,j}}{\Delta x} , \qquad \text{(F.13a)}$$

Table F.2: Minimum and maximum values after one rotation of the cone using all the methods.

Scheme	Minimum	Maximum
Split Godunov	0.0000	0.3300
Unsplit Godunov	0.0000	0.3247
CTU Godunov	0.0000	0.3299
Lax-Wendroff	-0.7970	0.8486
Split HOG	0.0000	0.8604
Unsplit HOG	0.0000	0.8638
CTU HOG/Godunov	-0.0120	0.8575
CTU HOG	-0.0190	0.8589
Runge-Kutta HOG	0.0000	0.8697
Hancock-van Leer HOG	-0.0062	0.8529

Table F.3: Minimum and maximum values after one rotation of the slotted cylinder using all the methods.

Scheme	Minimum	Maximum
Split Godunov	0.0000	0.5883
Unsplit Godunov	0.0000	0.5794
CTU Godunov	0.0000	0.5882
Lax-Wendroff	-0.7945	1.2627
Split HOG	0.0000	0.9993
Unsplit HOG	-0.0005	0.9996
CTU HOG/Godunov	-0.0555	1.0625
CTU HOG	-0.0585	1.0736
Runge-Kutta HOG	0.0000	0.9999
Hancock-van Leer HOG	-0.0332	0.9985

and in the y-direction

$$P_{i,j}(y) = u_{i,j} + \widetilde{\Delta_y u}\frac{y - y_{i,j}}{\Delta y} . \qquad \text{(F.13b)}$$

The terms $\widetilde{\Delta_x u}$ and $\widetilde{\Delta_y u}$ are defined by limiters (see Chapter 8).

From the above methods I may get second-order time accuracy by defining the time-centered, cell-edge values as

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j,l} = u^n_{i,j} + \frac{1}{2}(1 - \eta_x)\,\widetilde{\Delta_x u} , \qquad \text{(F.14a)}$$

and

$$u^{n+\frac{1}{2}}_{i-\frac{1}{2},j,r} = u^n_{i,j} - \frac{1}{2}(1 + \eta_x)\,\widetilde{\Delta_x u} , \qquad \text{(F.14b)}$$

where $\eta_x = a\Delta t/\Delta x$. The terms η_x and η_y are signed versions of ν_x and ν_y. Similar definitions are used for the cell edges in the y-direction.

Now I explore how I extend these one-dimensional methods to two space dimensions.

F.5.2 High-Resolution Methods in Multiple Spatial Dimensions

The first three ways to extend schemes to multiple spatial dimensions are simply extensions of the methods used for the first-order Godunov schemes. The operator split and unsplit methods are extremely straightforward, but the CTU scheme is worth exploring.

To get second-order accuracy I use a Taylor expansion for each time-centered cell-edge value

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j,l} = u^n_{i,j} + \frac{\Delta t}{2}\frac{\partial u}{\partial t} + \frac{\Delta x}{2}\frac{\partial u}{\partial x} , \qquad \text{(F.15a)}$$

and

$$u^{n+\frac{1}{2}}_{i,j+\frac{1}{2},b} = u^n_{i,j} + \frac{\Delta t}{2}\frac{\partial u}{\partial t} + \frac{\Delta y}{2}\frac{\partial u}{\partial y} . \qquad \text{(F.15b)}$$

I can replace $\partial u/\partial t$ with $-\partial f/\partial x - \partial g/\partial y$ in a manner similar to the derivation of the Lax-Wendroff method. This gives

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j,l} = u^n_{i,j} - \frac{\Delta t}{2}\left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}\right) + \frac{\Delta x}{2}\frac{\partial u}{\partial x} , \qquad \text{(F.16a)}$$

and

$$u^{n+\frac{1}{2}}_{i,j+\frac{1}{2},b} = u^n_{i,j} - \frac{\Delta t}{2}\left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}\right) + \frac{\Delta y}{2}\frac{\partial u}{\partial y} . \qquad \text{(F.16b)}$$

I use these expressions later in developing another method. Remembering that $f = au$

and $g = bu$, then gathering like terms results in

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j,l} = u^n_{i,j} + \frac{1}{2}(\Delta x - \Delta t a)\frac{\partial u}{\partial x} - \frac{\Delta t}{2}\frac{\partial g}{\partial y} . \qquad \text{(F.17a)}$$

and

$$u^{n+\frac{1}{2}}_{i,j+\frac{1}{2},b} = u^n_{i,j} + \frac{1}{2}(\Delta y - \Delta t b)\frac{\partial u}{\partial y} - \frac{\Delta t}{2}\frac{\partial f}{\partial x} . \qquad \text{(F.17b)}$$

Evaluated numerically the above expressions become

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j,l} = u^n_{i,j} + \frac{1}{2}(\Delta x - \Delta t a)\frac{\tilde{\Delta}_x u}{\Delta x} - \frac{\sigma_y}{2}\left(g^n_{i,j+\frac{1}{2},M} - g^n_{i,j-\frac{1}{2},M}\right) . \qquad \text{(F.18a)}$$

and

$$u^{n+\frac{1}{2}}_{i,j+\frac{1}{2},b} = u^n_{i,j} + \frac{1}{2}(\Delta y - \Delta t b)\frac{\tilde{\Delta}_y u}{\Delta y} - \frac{\sigma_x}{2}\left(f^n_{i+\frac{1}{2},j,R} - f^n_{i-\frac{1}{2},j,R}\right) . \qquad \text{(F.18b)}$$

The original CTU method presented above used the last terms in each of the last two expressions in defining the time-centered cell-edge values used in (F.10c). Applying the HOG polynomial reconstruction given in the previous section provides values for the new terms in the expansions. Two separate methods arise from this derivation: I get second-order accuracy with Godunov fluxes being used as with the first-order CTU method or I may use second-order fluxes in the place of the first- order fluxes. The first of these two methods I call the CTU HOG/Godunov method and the second CTU/HOG.

In [159, 158], an alternate method for extending HOG methods to second-order time accuracy was presented. This method was developed in one dimension and is similar in flavor to the two-step Lax-Wendroff method. Using the above stated derivation I can extend this method to two (or more) dimensions. I substitute numerical approximations directly into (F.16a) and (F.16b). This gives expressions for the time-centered cell-edge values of

$$u^{n+\frac{1}{2}}_{i+\frac{1}{2},j,l} = u^n_{i+\frac{1}{2},j,l} - \frac{\sigma_x}{2}\left(f^n_{i+\frac{1}{2},j,l} - f^n_{i-\frac{1}{2},j,r}\right) - \frac{\sigma_y}{2}\left(f^n_{i,j+\frac{1}{2},b} - f^n_{i,j-\frac{1}{2},t}\right) , \qquad \text{(F.19a)}$$

$$u^{n+\frac{1}{2}}_{i-\frac{1}{2},j,r} = u^n_{i-\frac{1}{2},j,r} - \frac{\sigma_x}{2}\left(f^n_{i+\frac{1}{2},j,l} - f^n_{i-\frac{1}{2},j,r}\right) - \frac{\sigma_y}{2}\left(f^n_{i,j+\frac{1}{2},b} - f^n_{i,j-\frac{1}{2},t}\right) . \qquad \text{(F.19b)}$$

$$u^{n+\frac{1}{2}}_{i,j+\frac{1}{2},b} = u^n_{i,j+\frac{1}{2},b} - \frac{\sigma_x}{2}\left(f^n_{i+\frac{1}{2},j,l} - f^n_{i-\frac{1}{2},j,r}\right) - \frac{\sigma_y}{2}\left(f^n_{i,j+\frac{1}{2},b} - f^n_{i,j-\frac{1}{2},t}\right) , \qquad \text{(F.19c)}$$

and

$$u^{n+\frac{1}{2}}_{i,j-\frac{1}{2},t} = u^n_{i,j-\frac{1}{2},t} - \frac{\sigma_x}{2}\left(f^n_{i+\frac{1}{2},j,l} - f^n_{i-\frac{1}{2},j,r}\right) - \frac{\sigma_y}{2}\left(f^n_{i,j+\frac{1}{2},b} - f^n_{i,j-\frac{1}{2},t}\right) . \qquad \text{(F.19d)}$$

These estimates differ from previous schemes by not requiring Riemann solvers. Then

(F.10c) is used to update the grid values. This is referred to as the Hancock-van Leer HOG method. Unlike the CTU schemes, the CFL limit for this scheme is given by

$$\nu_x + \nu_y \leq 1 .$$

This is because cell-to-cell interactions are ignored in the predictor step.

The next method I study here is a TVD Runge-Kutta method introduced by Shu [169, 65, 66]. These methods were shown to be TVD when the coefficients of the time discretization meet certain conditions. These multistage algorithms are written in the following form

$$u^i = \sum_{k=0}^{i-1} \left[\alpha_{ik} u^k + \beta_{ik} \Delta t L\left(u^k\right) \right] , \tag{F.20a}$$

where the semi-discrete differential operator is defined by

$$\frac{\partial u}{\partial t} = L(u) , \tag{F.20b}$$

and α_{ik} and β_{ik} are coefficients. The criteria for this to produce TVD results given an appropriate spatial operator is a CFL condition

$$\nu \leq \frac{\alpha_{ik}}{|\beta_{ik}|} . \tag{F.20c}$$

where

$$\nu_x + \nu_y \leq \quad .$$

If β_{ik} is negative, the spatial operator must be antiupwind [65, 160]. A number of schemes can be defined with the second- and third-order methods being particularly useful. The second-order method turns out to be the classic modified Euler or Heun scheme

$$u^1_{i,j} = u^n_{i,j} + \Delta t L(u^n) , \tag{F.21a}$$

and

$$u^{n+1}_{i,j} = u_{i,j} + \frac{\Delta t}{2} \left[L(u^n) + L\left(u^1\right) \right] , \tag{F.21b}$$

with a CFL condition of $\nu \leq 1$. It is notable that Riemann solvers are needed at each step of the multistep integration.

For convenience, the CFL limits for the schemes studied in this appendix are given in Table F.4

F.6 Results for the Second-Order Methods

This section shows and discusses solutions to the test problems by the second-order methods described above. Before continuing to this, I show the results that a classic

Table F.1: CFL limits for all the methods.

Scheme	Limit
Split Godunov	$\max(\nu_x, \nu_y) \leq 1$
Unsplit Godunov	$\nu_x + \nu_y \leq 1$
CTU Godunov	$\max(\nu_x, \nu_y) \leq 1$
Lax-Wendroff	$\max(\nu_x, \nu_y) \leq 1$
Split HOG	$\max(\nu_x, \nu_y) \leq 1$
Unsplit HOG	$\nu_x + \nu_y \leq 1$
CTU HOG/Godunov	$\max(\nu_x, \nu_y) \leq 1$
CTU HOG	$\max(\nu_x, \nu_y) \leq 1$
Runge-Kutta HOG	$\nu_x + \nu_y \leq 1$
Hancock-van Leer HOG	$\nu_x + \nu_y \leq 1$

second-order method produces. Figures F.10 and F.11 show the operator split Lax-Wendroff solutions to the test problems. Both of these solutions are unacceptable. The large error near the lower boundary is the consequence of boundary conditions. The boundaries are set to a symmetrical condition which does not damp out errors at the boundary. Eventually, the solution undergoes boundless growth because of this. If the solutions are set to zero at the boundary (errors flow out of the domain), the solutions remain bounded. Therefore, the boundary conditions used here represent a worse case analogous to reflective boundary conditions in fluid flow simulations.

Table F.1 shows the economy of each scheme. All are more expensive than the Lax-Wendroff method, with the split and unsplit HOG methods being the least expensive followed by the Runge-Kutta HOG method. The CTU and Hancock-van Leer methods are all very expensive. The bulk of this expense seems to be related to memory access time, which favors the Runge-Kutta type method. In terms of economy, the more classical split method appears to be the winner.

In all the HOG-type methods shown here, the superbee limiter is used to give the highest resolution possible. Other limiters are briefly discussed later in the appendix. The split HOG method gives excellent results in terms of resolution and solution symmetry (see Fig. F.12). The bridge in the slotted cylinder is only slightly eroded as shown by Fig. F.13.

The unsplit HOG method gives poor results in terms of solution symmetry and resolution as shown in Figs. F.14 and F.15. The problem with the unsplit method is

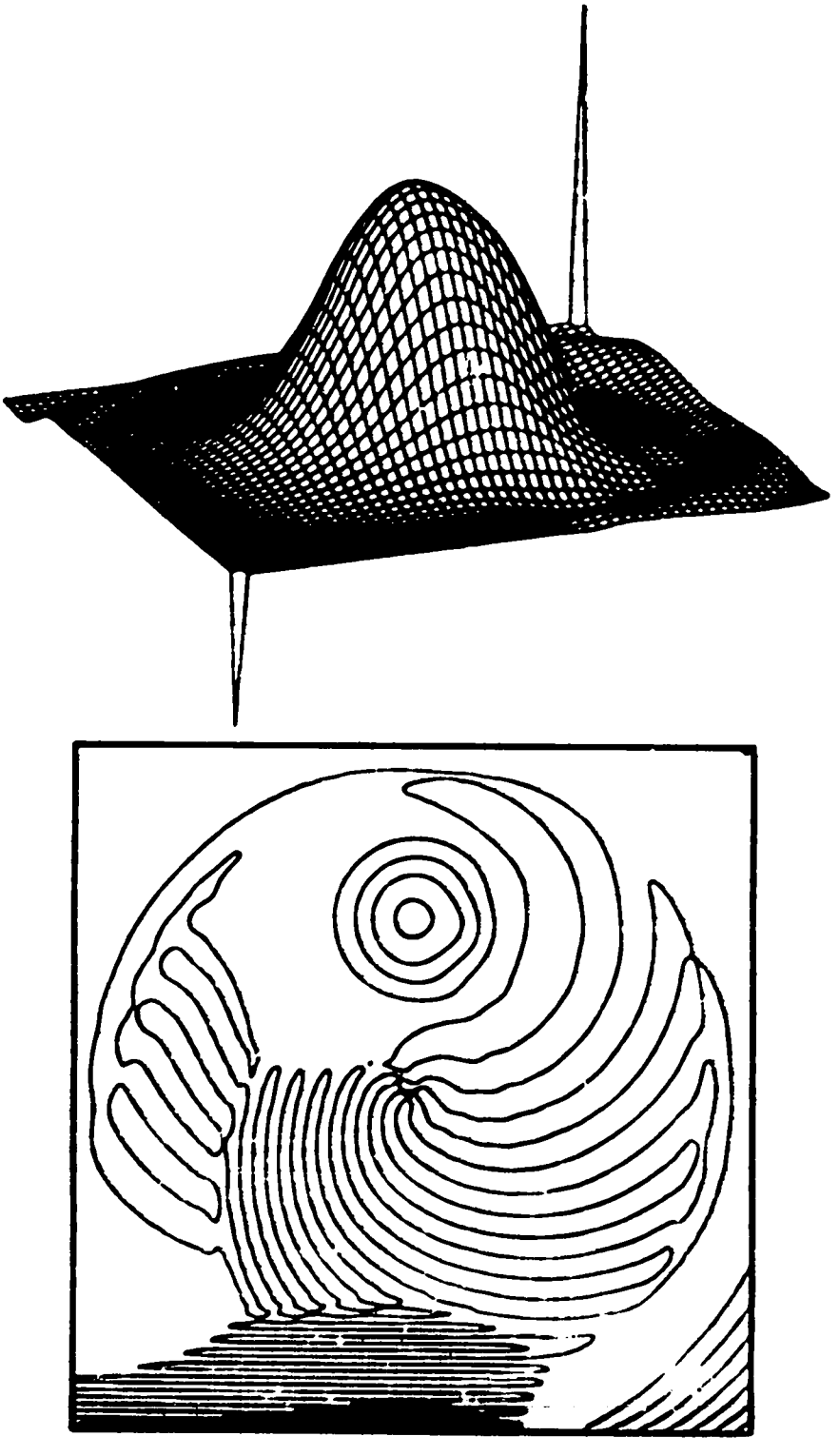


Figure F.10: The Lax-Wendroff method solution for the rotating cone shows the excessive dispersion errors of this method.

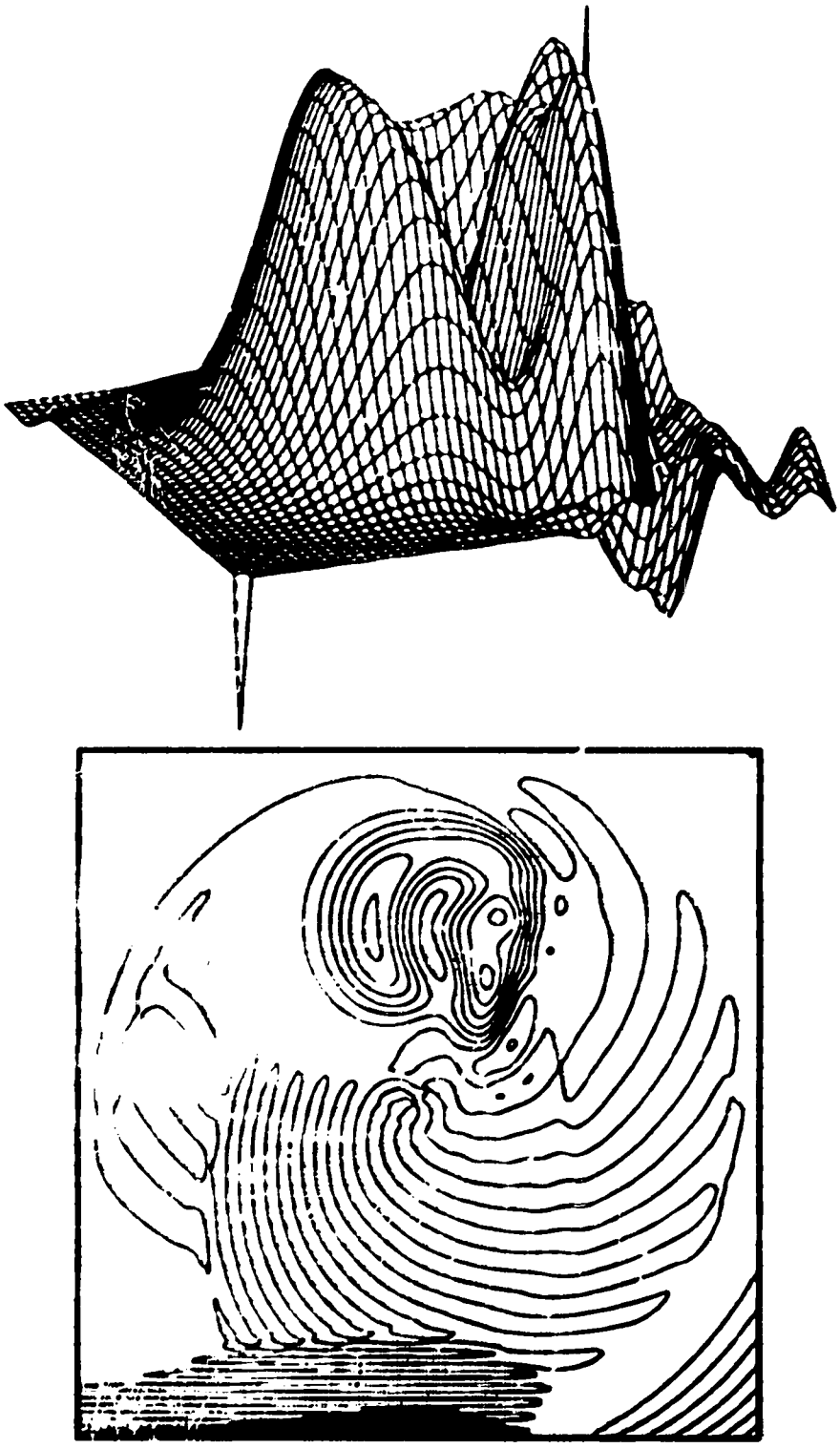


Figure F.11: The Lax-Wendroff method solution for the rotating slotted cylinder shows the excessive dispersion errors of this method.

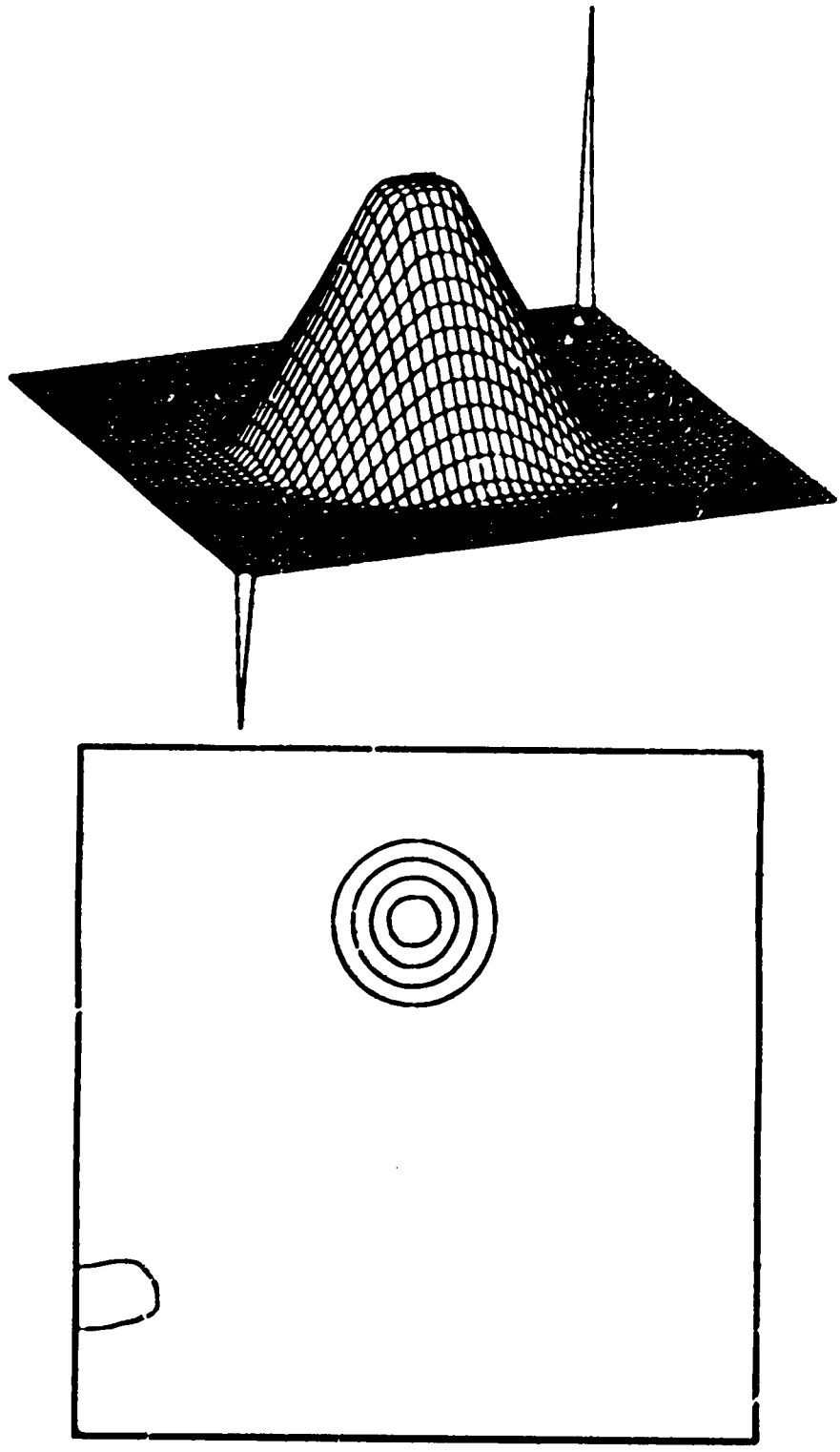


Figure F.12: The split HOG method solution for the rotating cone shows the high quality of this method.

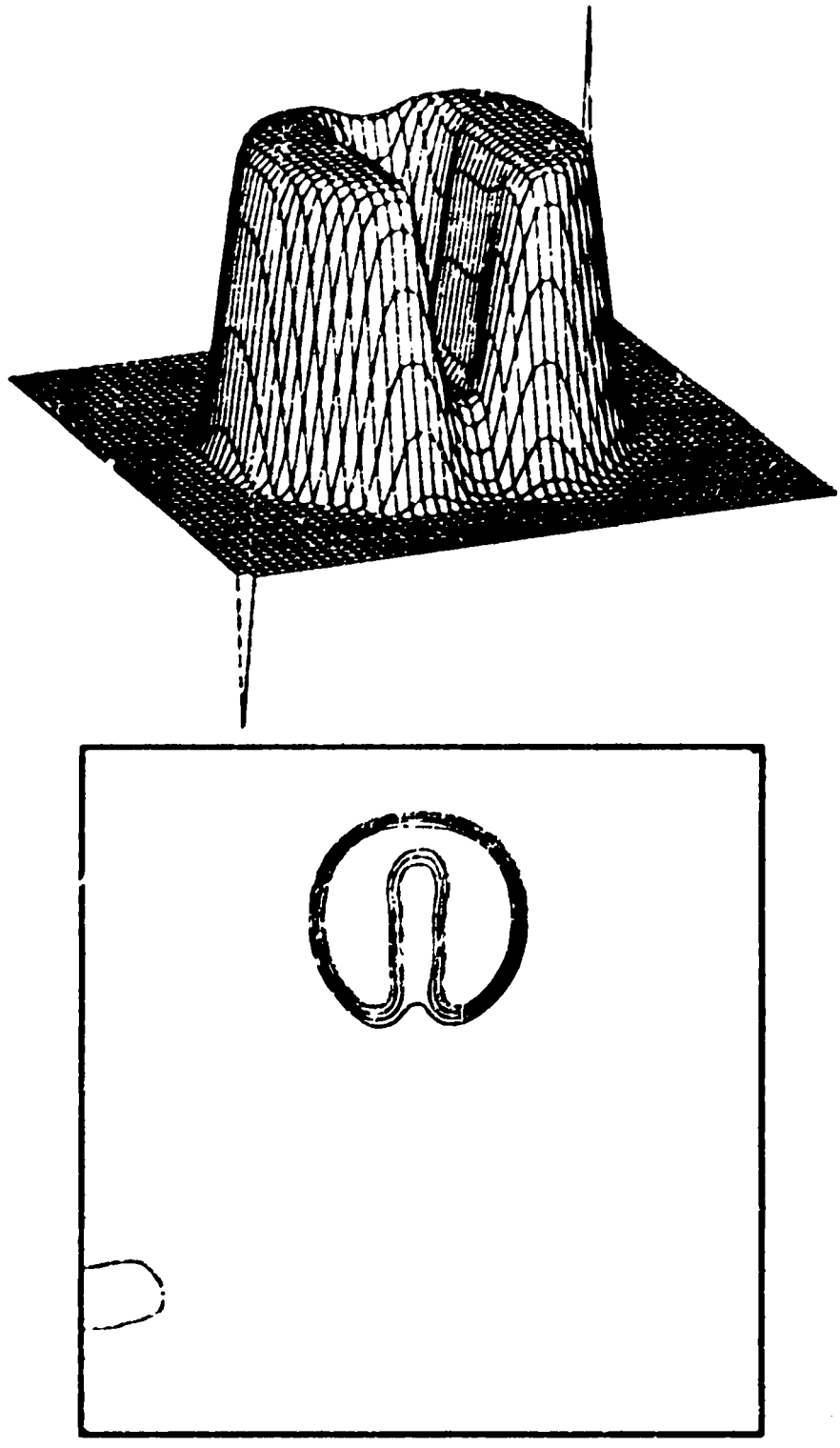


Figure F.13: The split HOG method solution for the rotating slotted cylinder shows the high quality of this method

Table F.5: Minimum and maximum values after one rotation of the cone for various limiters using the Runge-Kutta HOG method.

Limiter	Minimum	Maximum
Minmod	0.0000	0.6703
van Leer	0.0000	0.7754
Central	0.0000	0.8154
Superbee	0.0000	0.8697
Generalized Average n=2	-0.0277	0.8439

that the cross derivative terms ($\partial^2 u/\partial x \partial y$) are ignored. This problem has been noted by Sinolarkiewicz [242].

The solutions computed with the CTU Godunov/HOG and CTU HOG methods do not share this problem. Both methods have excellent symmetry qualities as shown by Figs. F.16 and F.18. The resolution is also quite high as can be seen in Figs. F.17 and F.19. These figures also show that the solutions are not monotone and also produce a great deal of high frequency but low amplitude noise. The solutions do not differ greatly as evidenced by the figures and the data in Tables F.2 and F.3, but the CTU HOG method is slightly noisier and less monotonic.

The Hancock-van Leer HOG method has many of the same characteristics as the CTU algorithm, but the oscillations are smaller and the actual resolution is improved. These two features are evident in Figs. F.20 and F.21. This method produces the best reproduction of the "bridge" in the slotted cylinder problem.

The Runge-Kutta HOG method improves on all these methods. As Figs. F.22 and F.23 demonstrate, the problems with the above methods are cured. The solutions is of slightly better quality than the split HOG method.

F.7 Test of Various Limiters

This section briefly discusses the performance of the HOG methods for different choices of flux limiters. Tables F.5 and F.6 show the minimum and maximum values for each of the limiter for the test problems. In all cases, the Runge-Kutta HOG method is used.

The figures that follow show that the choice of limiter can have a profound influence on the quality of the solution. The minmod limiter provides the lowest resolution second-order solution as is shown by Figs. F.24 and F.25. The van Leer and center limiters are somewhat better in resolution, but are still noticeably less resolved than

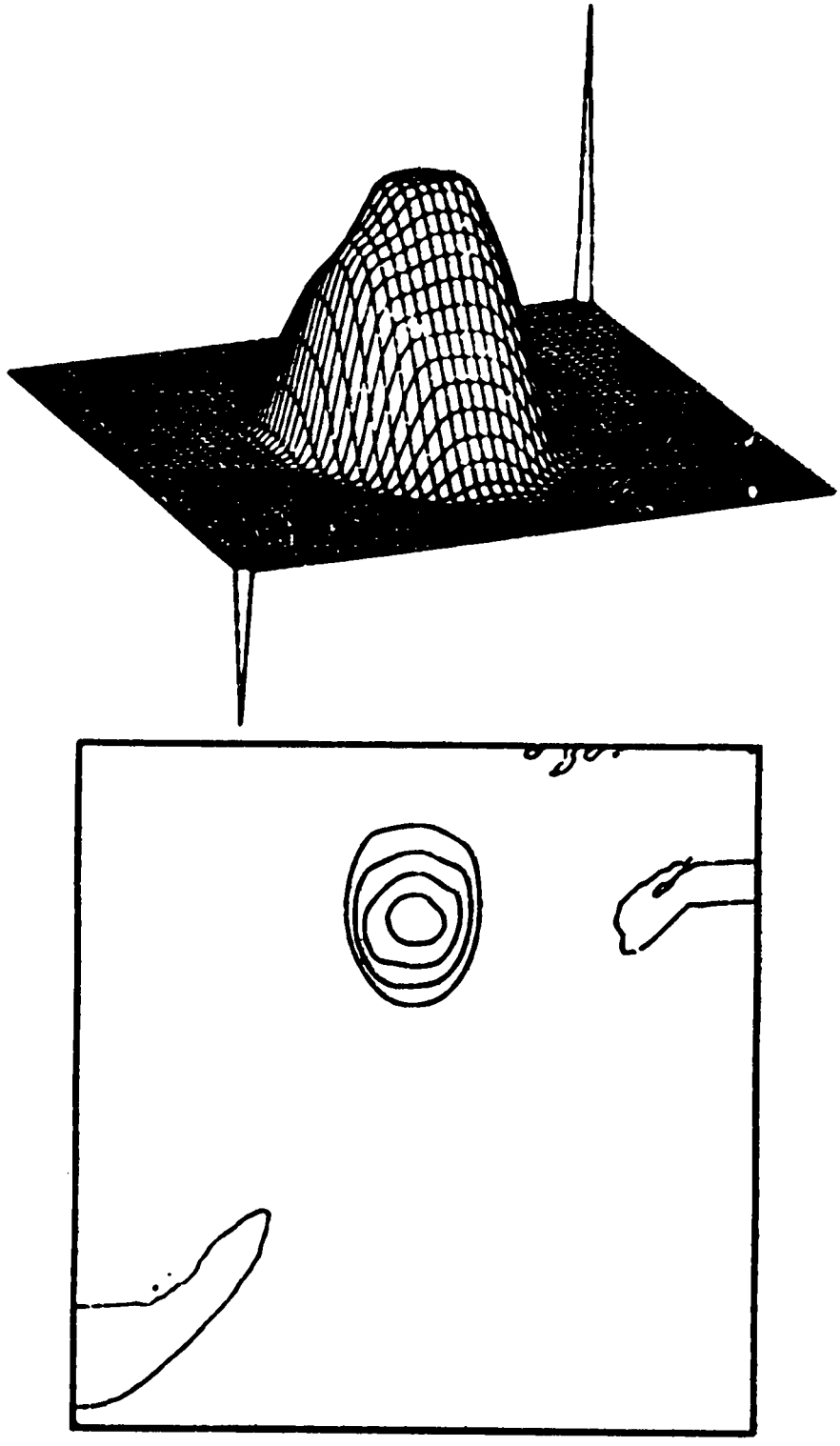


Figure F.14: The unsplit HOG method solution for the rotating cone shows the lack of symmetry of this method.

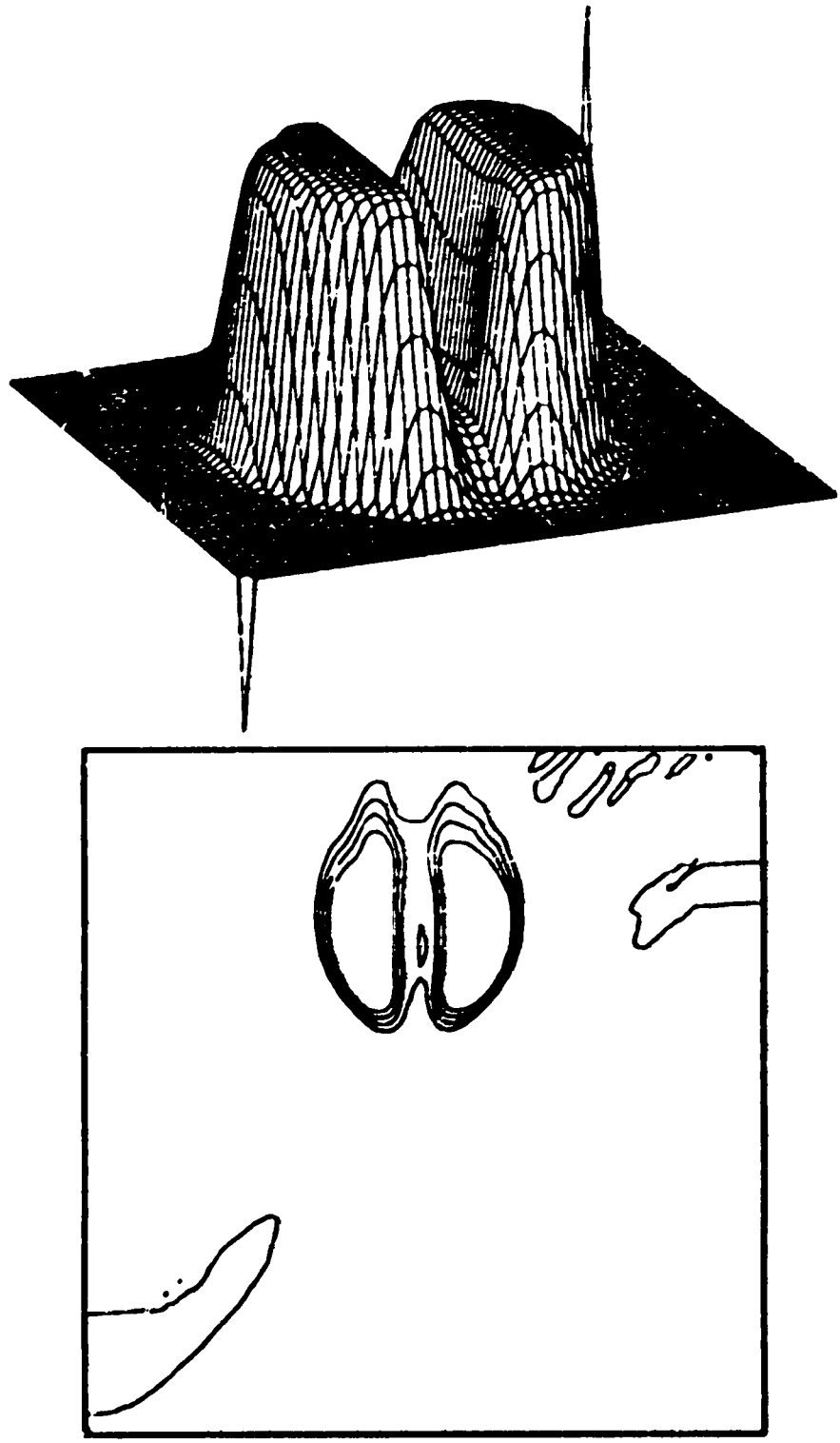


Figure F.15: The unsplit HOG method solution for the rotating slotted cylinder shows the lack of resolution of this method.

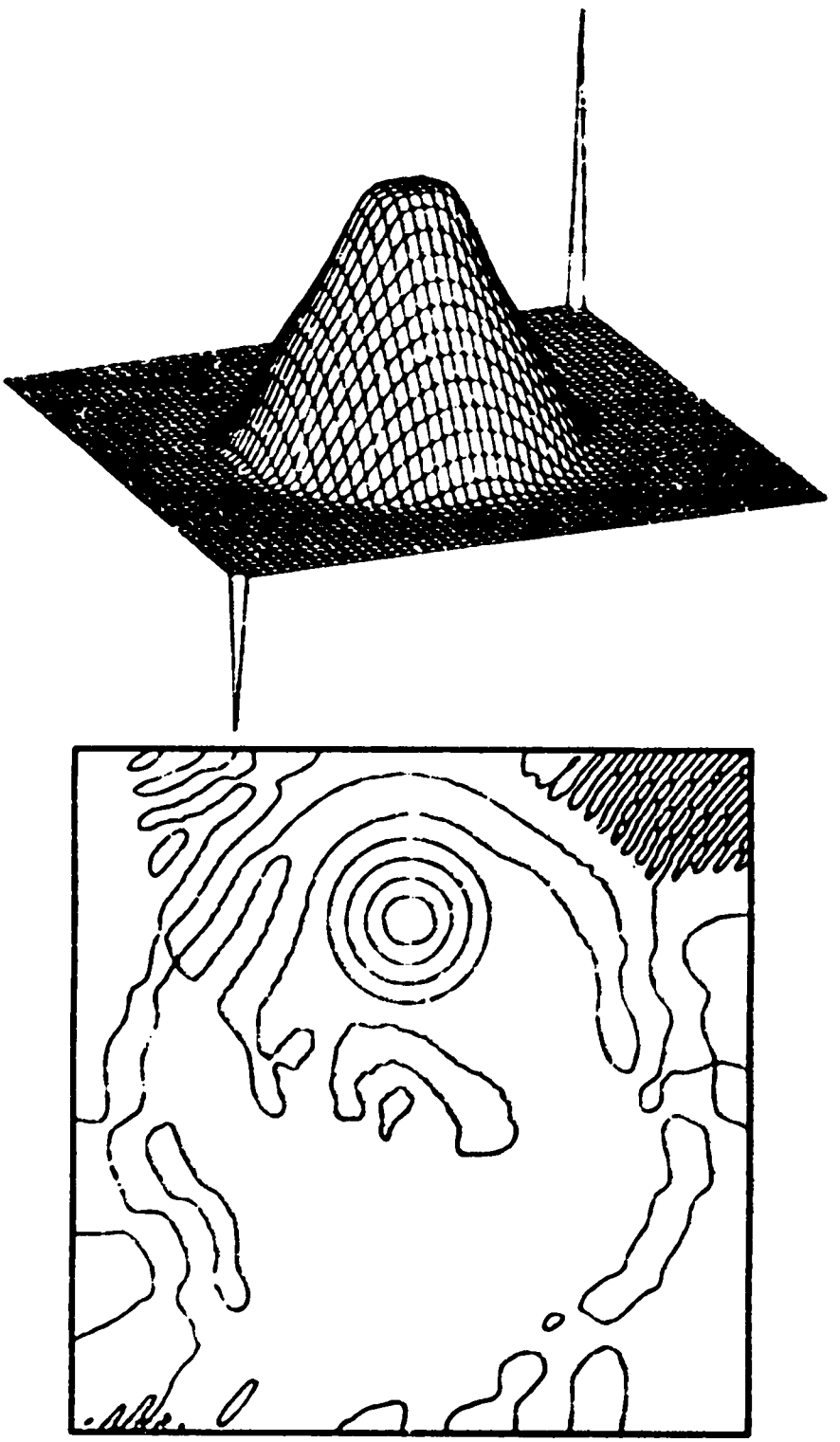


Figure F.16: The CTU Godunov/HOG method solution for the rotating cone shows the resolution and noise of this method.

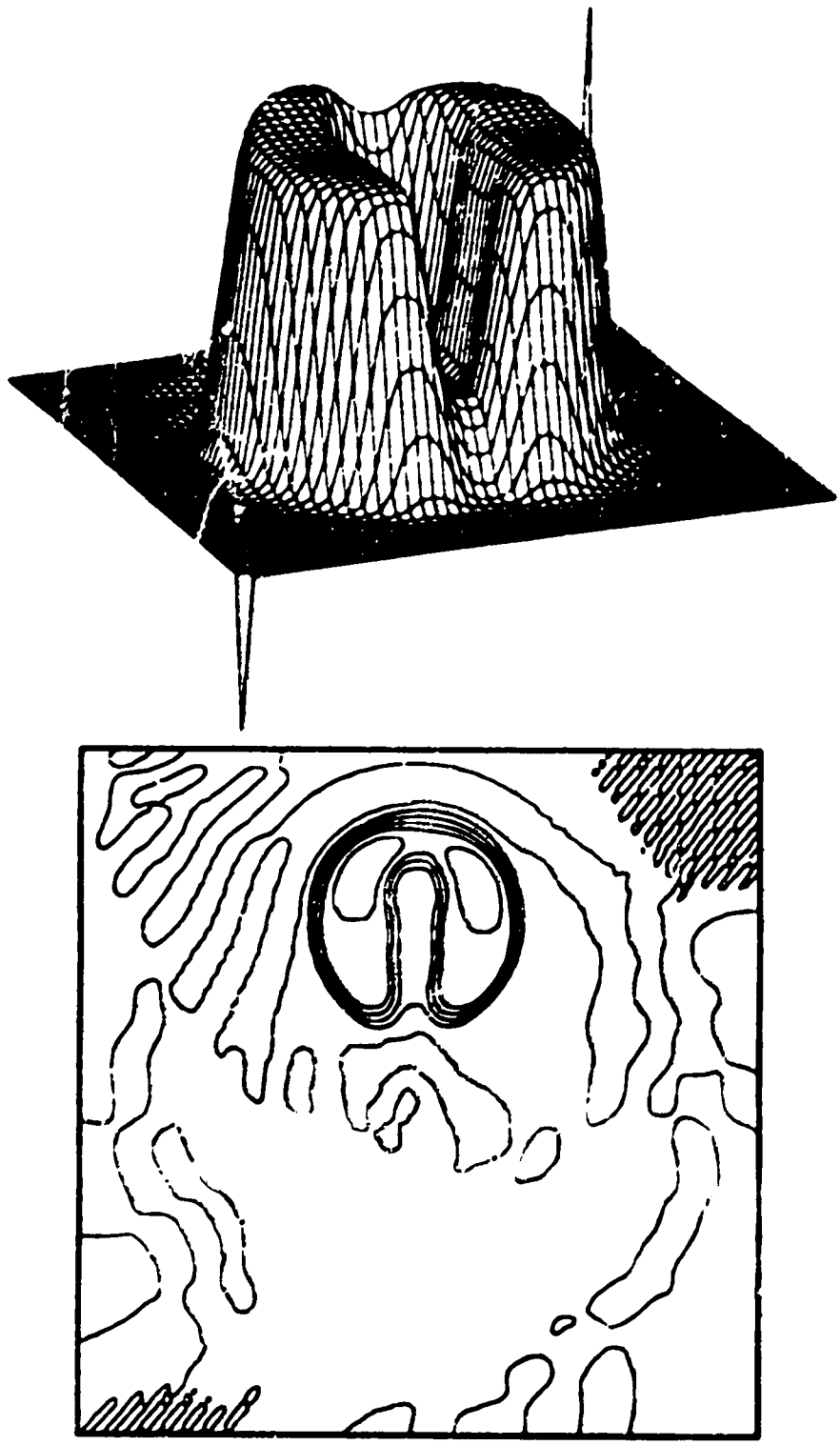


Figure F.17: The CTU Godunov/HOG method solution for the rotating slotted cylinder shows the resolution and noise of this method.

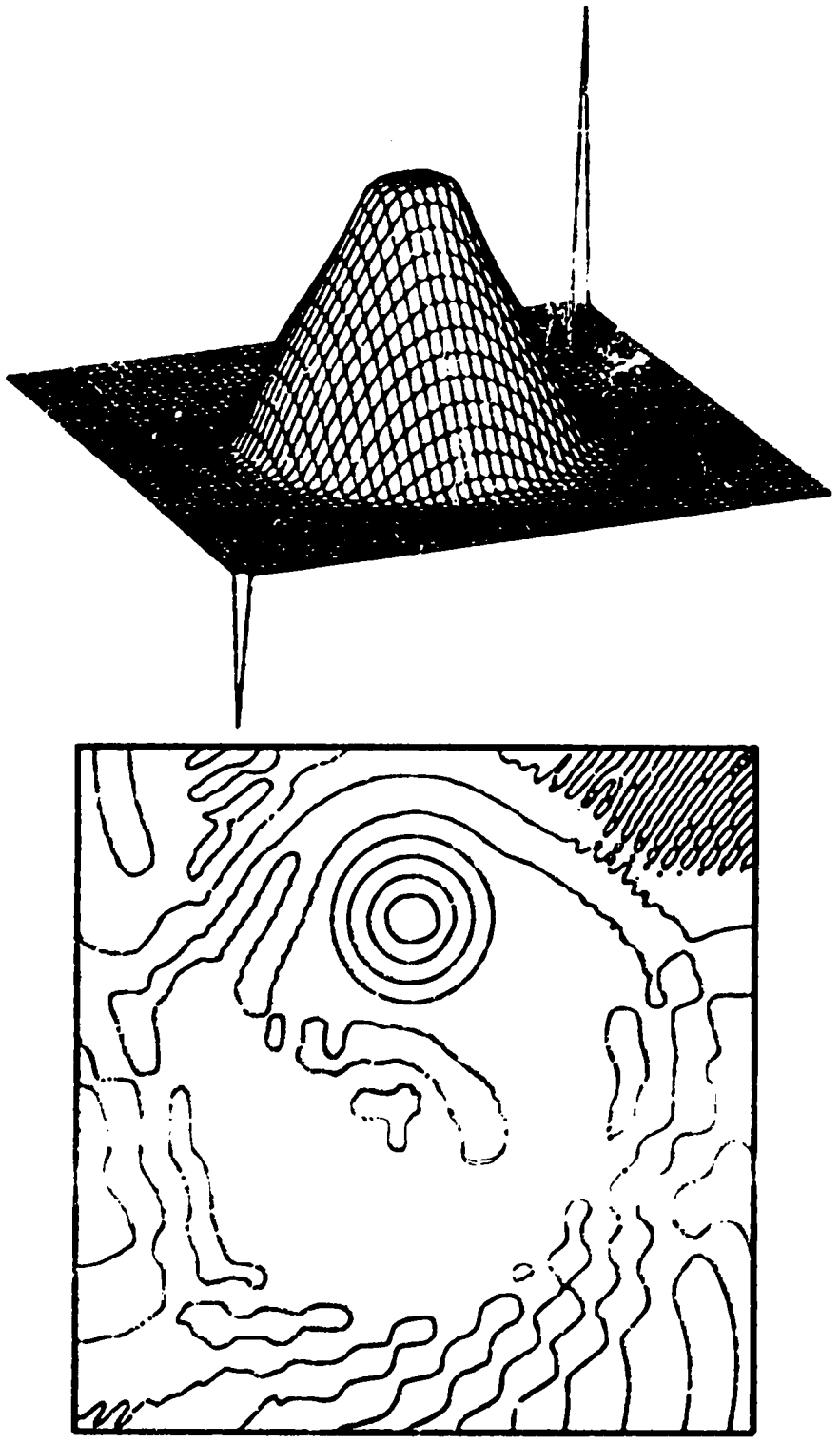


Figure F.18: The CTU HOG method solution for the rotating cone shows the resolution and noise of this method.

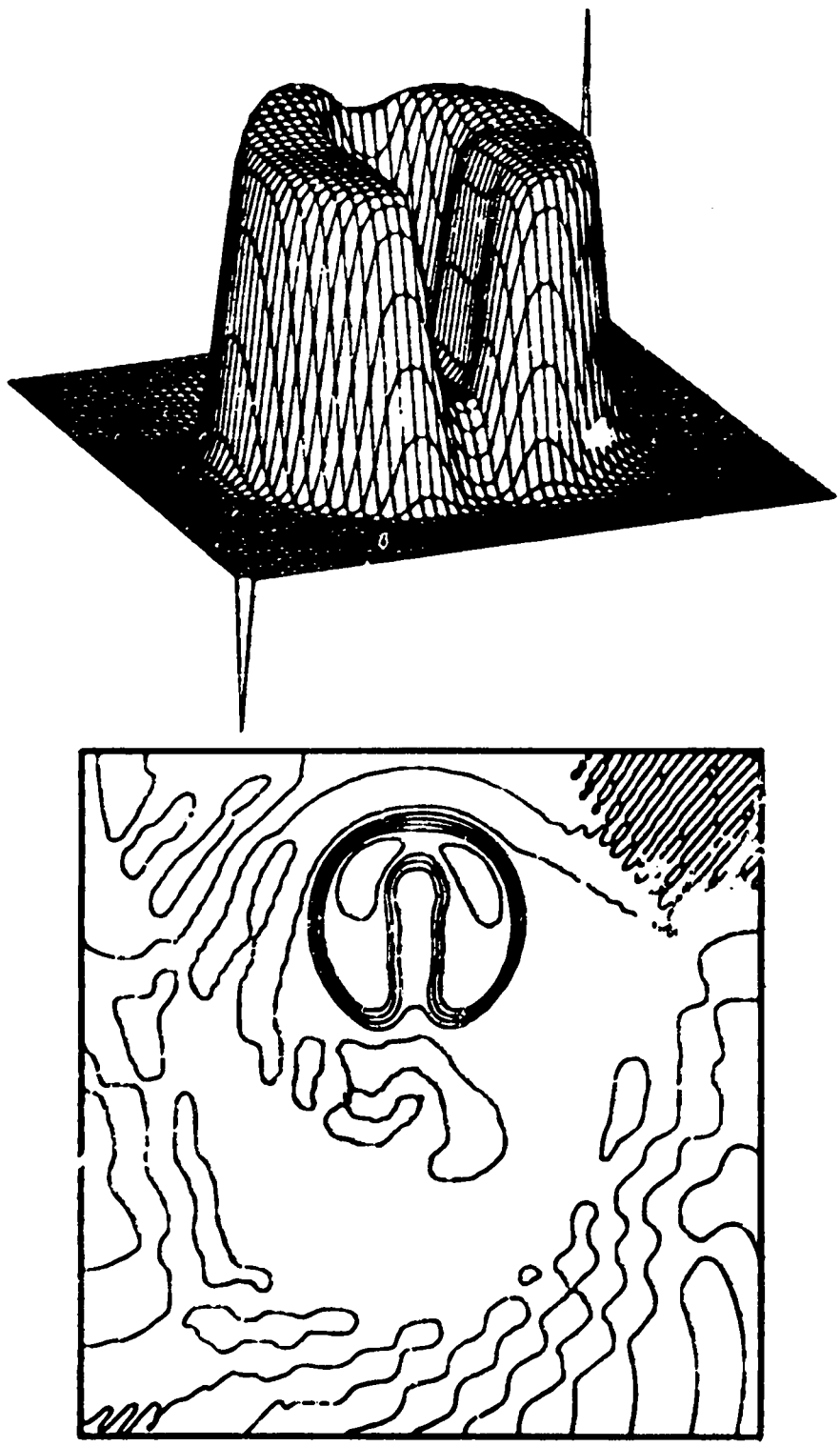


Figure F.19: The CTU HOG method solution for the rotating slotted cylinder shows the resolution and noise of this method.

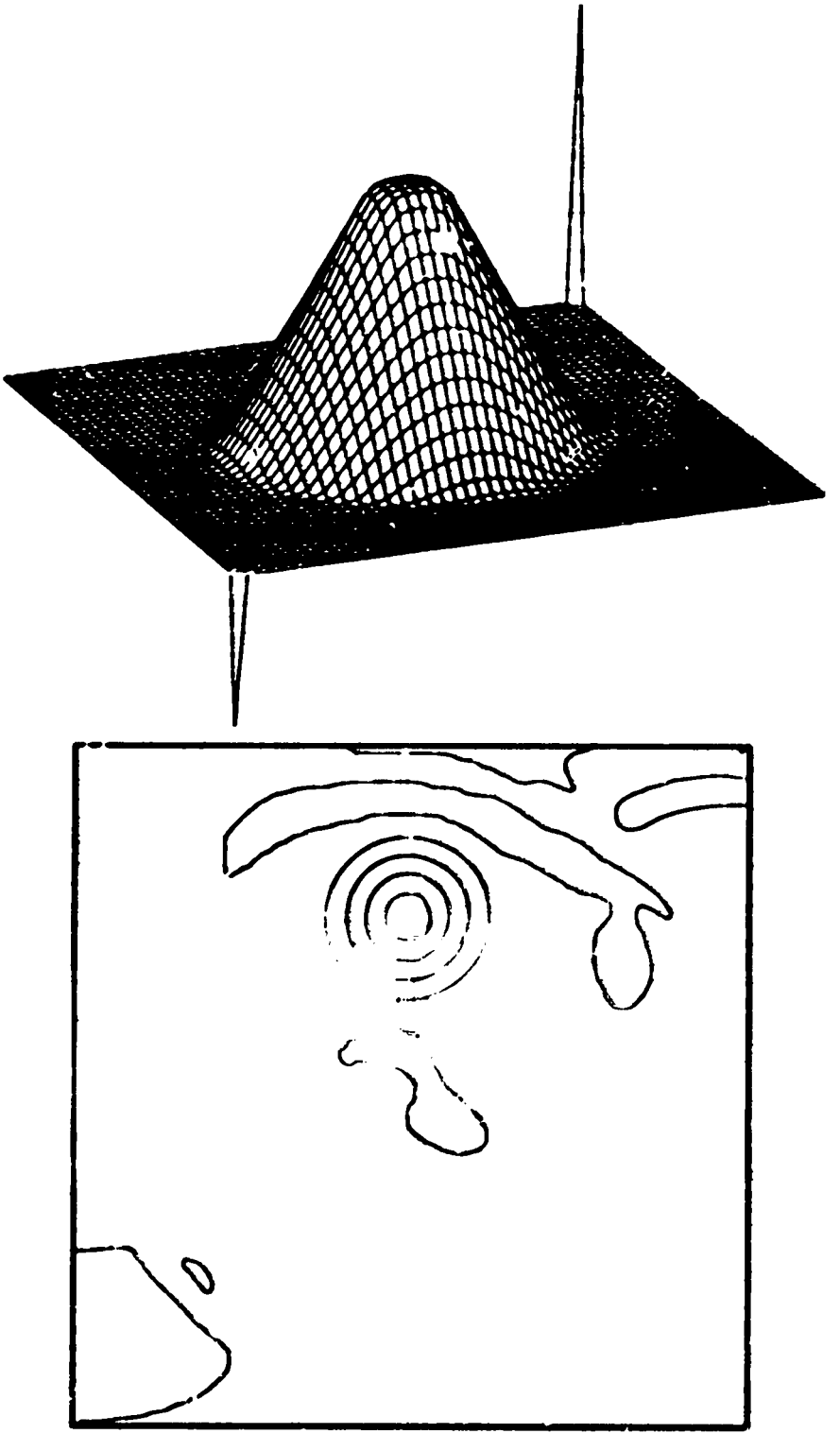


Figure F.20: The Hancock-van Leer HOC method solution for the rotating cone shows the resolution and reduced noise of this method

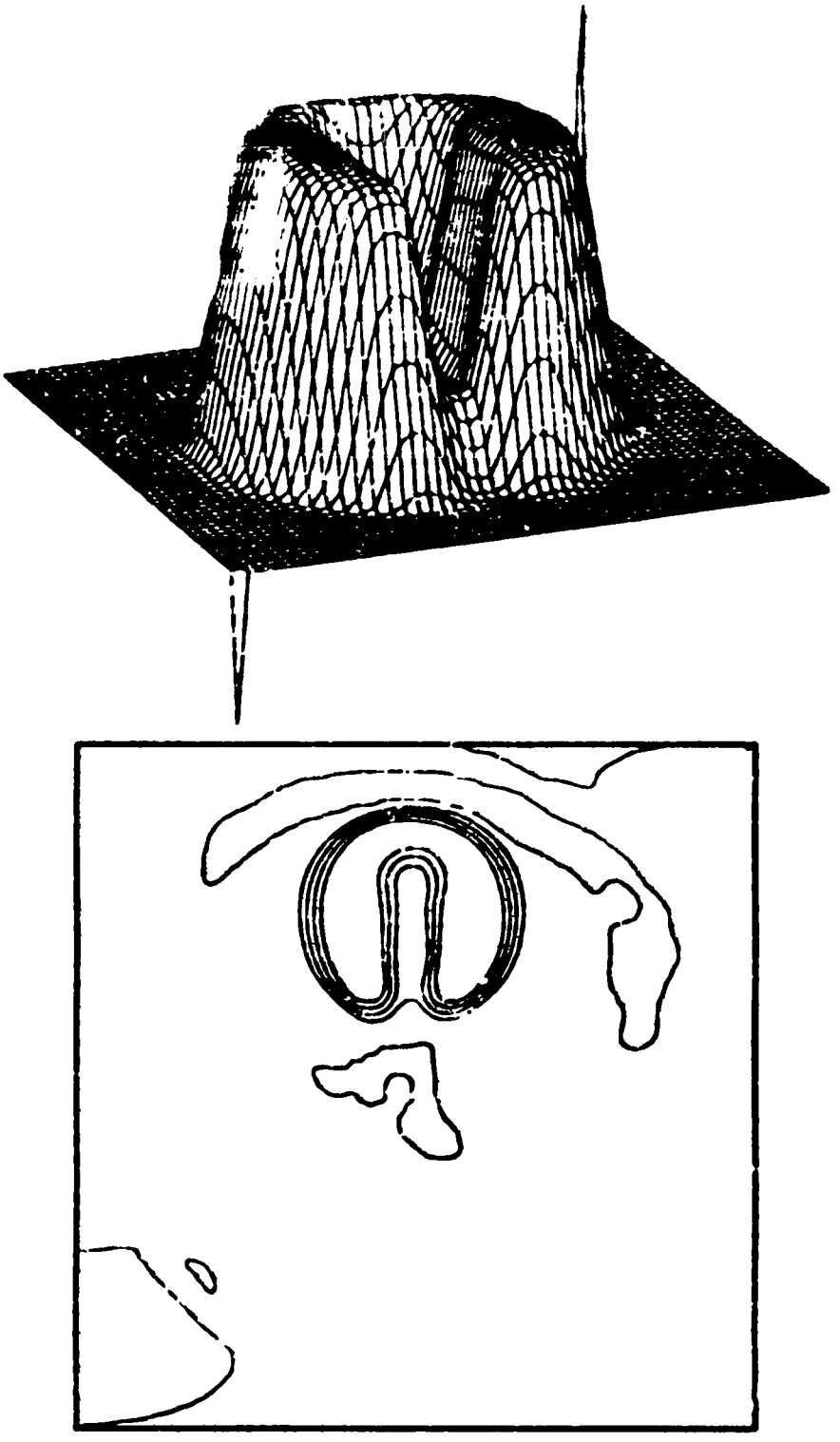


Figure F.21: The Hancock-van Leer HOG method solution for the rotating slotted cylinder shows the resolution and reduced noise of this method.

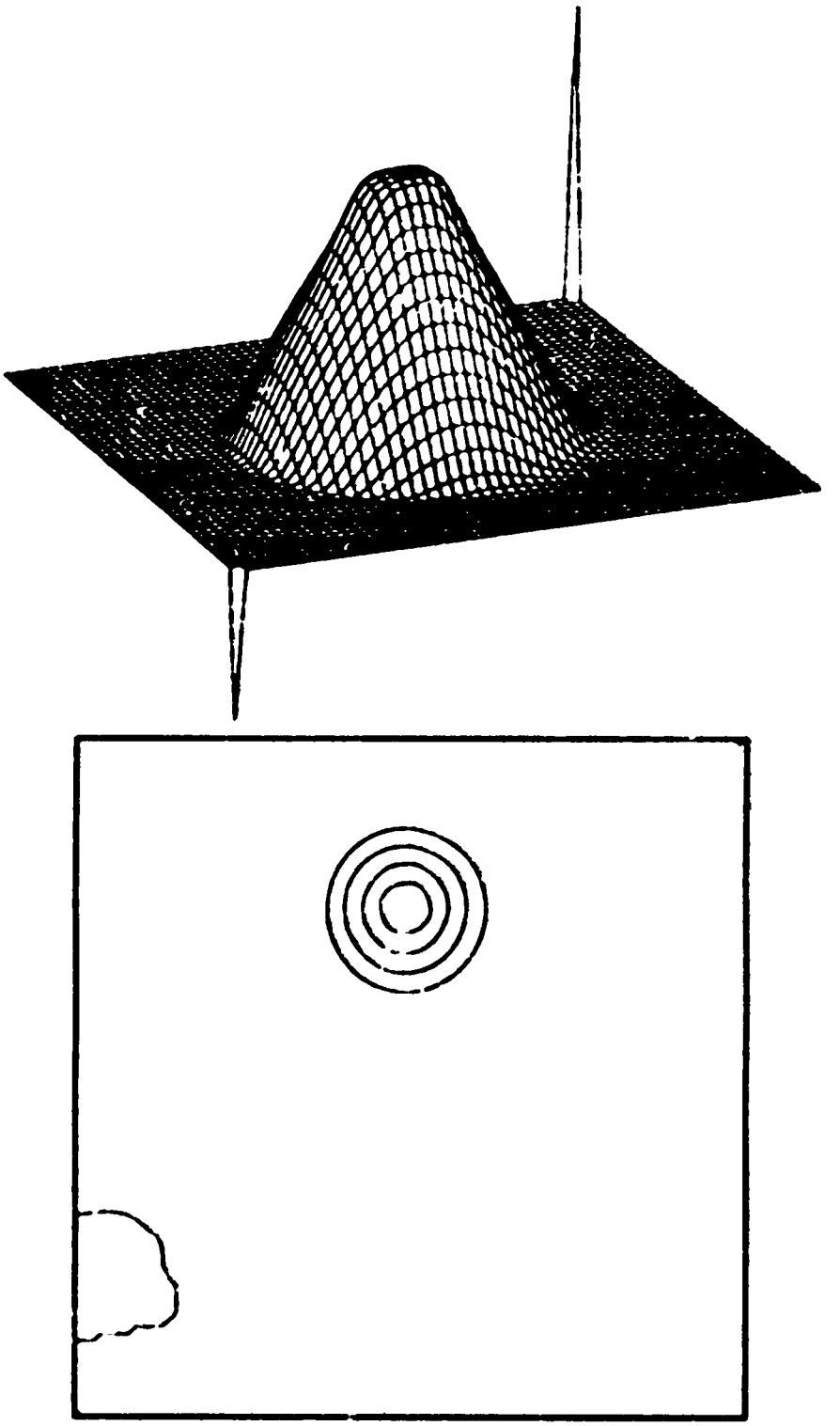


Figure F.22: The Runge-Kutta HOG method solution for the rotating cone shows the resolution and the lack of noise of this method.

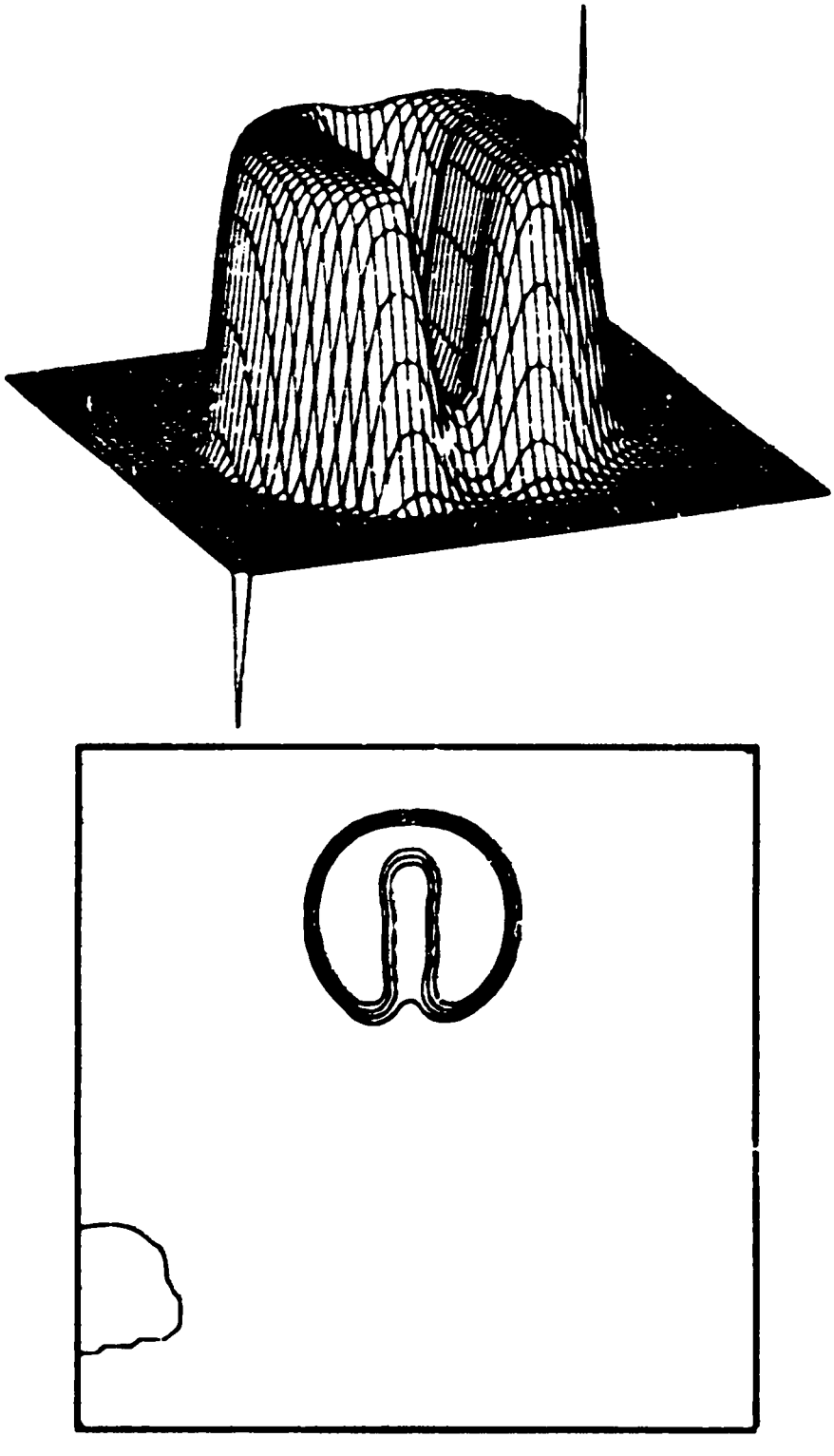


Figure F.23: The Runge-Kutta HOG method solution for the rotating slotted cylinder shows the resolution and the lack of noise of this method.

Table F.6: Minimum and maximum values after one rotation of the slotted cylinder for various limiters using the Runge-Kutta HOG method.

Limiter	Minimum	Maximum
Minmod	0.0000	0.7635
van Leer	0.0000	0.9237
Central	0.0000	0.9797
Superbee	0.0000	0.9999
Generalized Average n=2	-0.0759	1.0440

the superbee limiter. The center limiter solutions are given in Figs. F.26 and F.27 and the van Leer limiter solutions in Figs. F.28 and F.29. The generalized average limiter gives a more resolved solution, but at the cost of symmetry and monotonicity. These are shown in Figs. F.30 and F.31. The superbee limiter solutions were shown in Figs. F.22 and F.23.

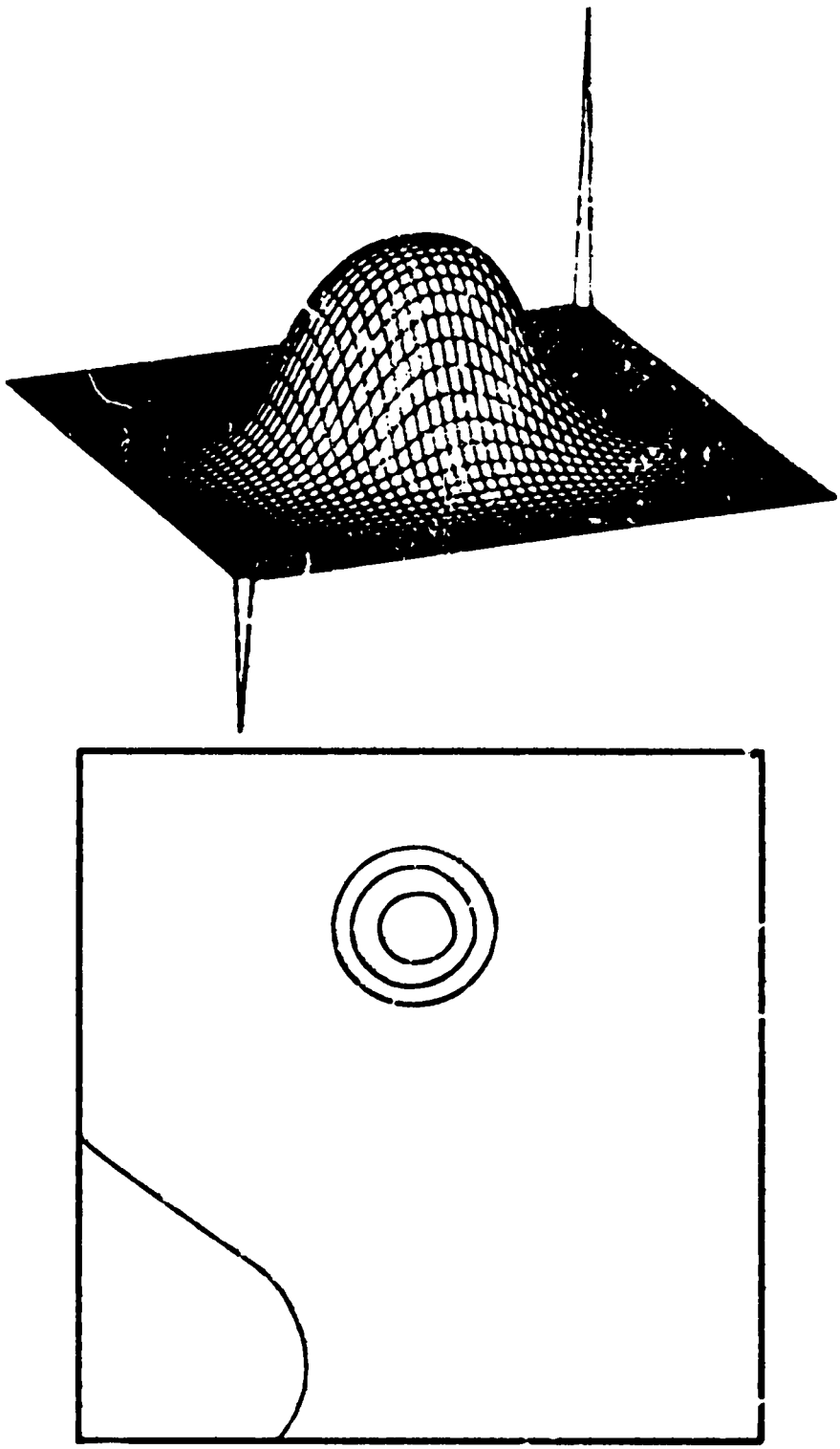


Figure F.24: The Runge-Kutta HOG method with the minmod limiter solution for the rotating cone shows the poor resolution of this limiter.

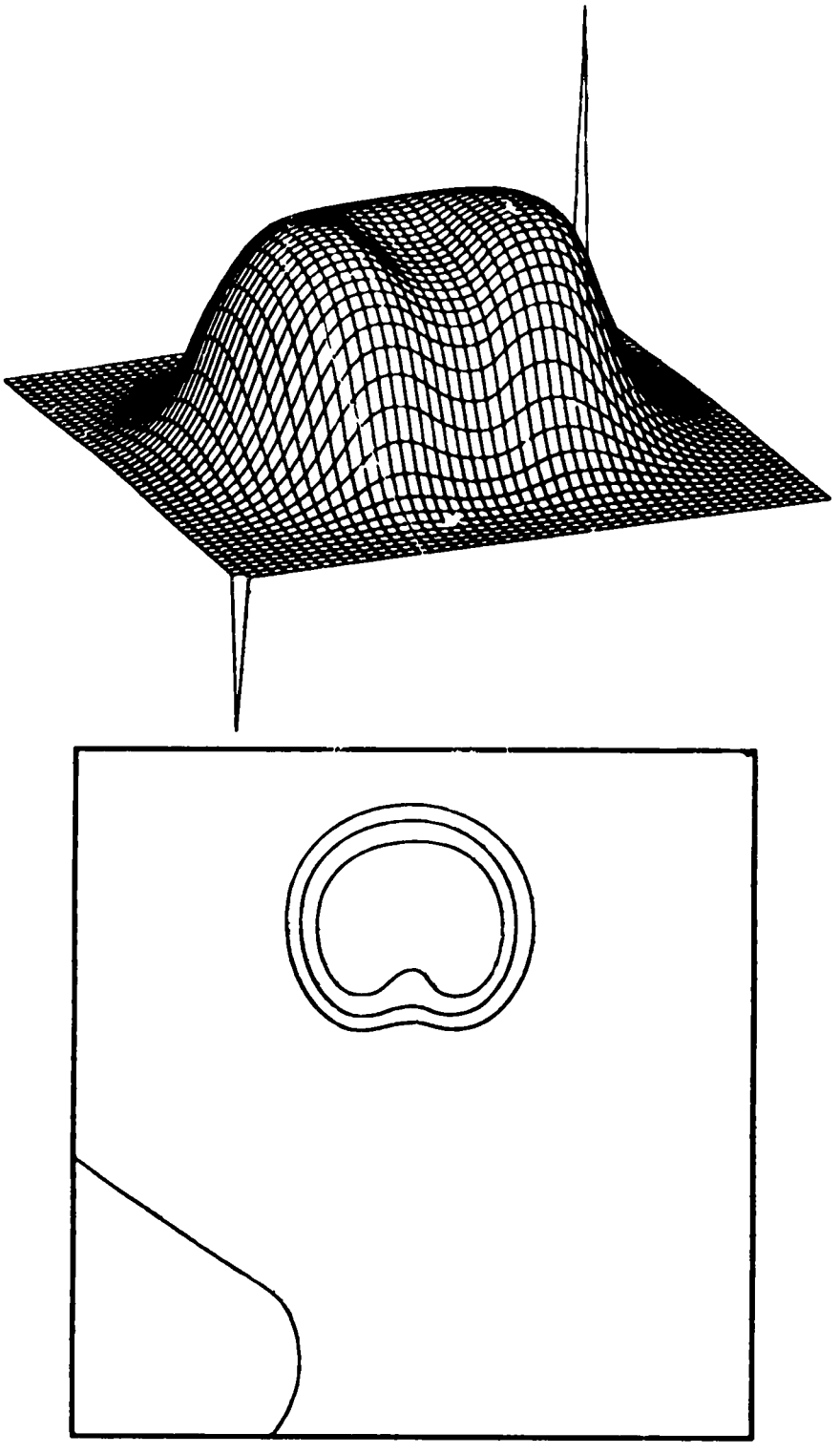


Figure F.25: The Runge-Kutta HOG method with the minmod limiter solution for the rotating slotted cylinder shows the poor resolution of this limiter.

F.8 Closing Remarks

Of the methods discussed in this chapter, the split HOG and Runge-Kutta HOG methods are the clear winners in terms of overall performance. The Runge-Kutta HOG methods are especially appealing because they can be extended to higher than second-order accuracy. This makes them important for consideration with ENO schemes or such schemes as the PPM [122]. The Hancock-van Leer method is an improvement in terms of performance and economy over the CTU-type methods. If a larger time step is desired, the split schemes seem to be quite effective. For systems of equations, this topic is in need of additional research. Split methods seem to have some intrinsic problems with systems [243]. Perhaps this swings the balance in favor of Runge-Kutta-type methods, but the performance of CTU-type methods also needs critical evaluation for systems.

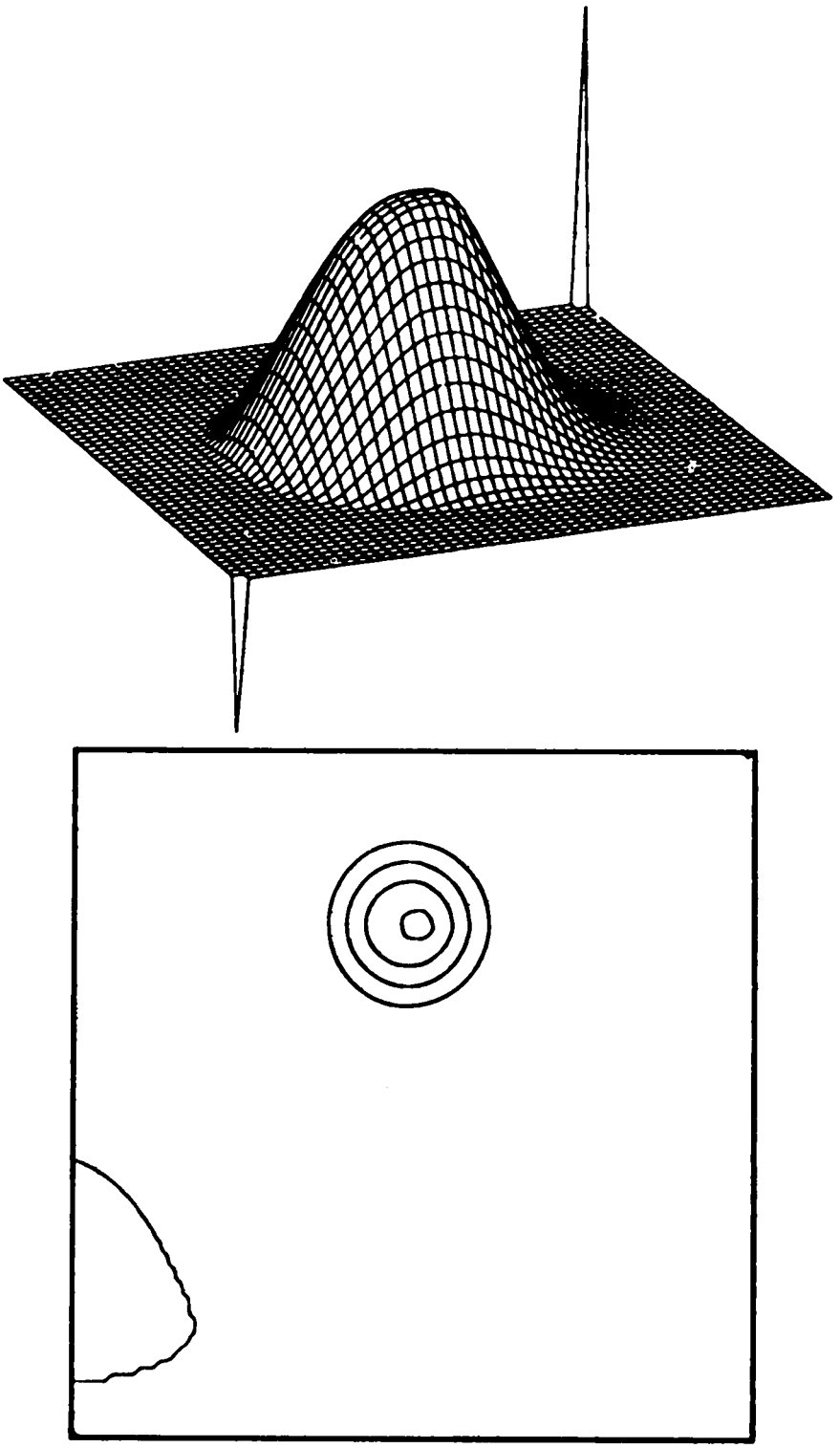


Figure F.26: The Runge-Kutta HOG method with the central limiter solution for the rotating cone shows the resolution of this limiter is nearly on par with the superbee limiter.

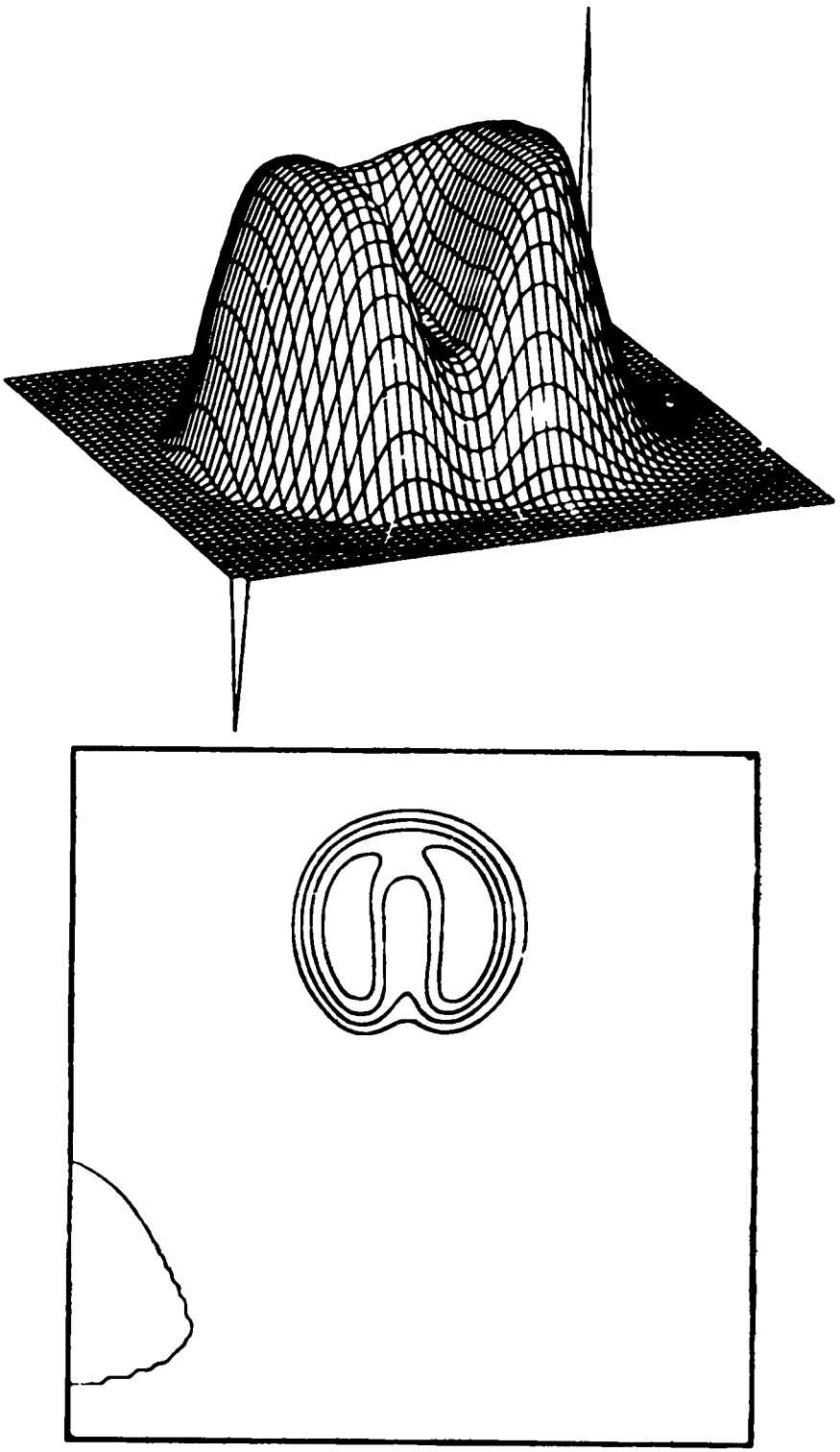


Figure F.27: The Runge-Kutta HOG method with the central limiter solution for the rotating slotted cylinder shows the resolution of this limiter is nearly on par with the superbee limiter.

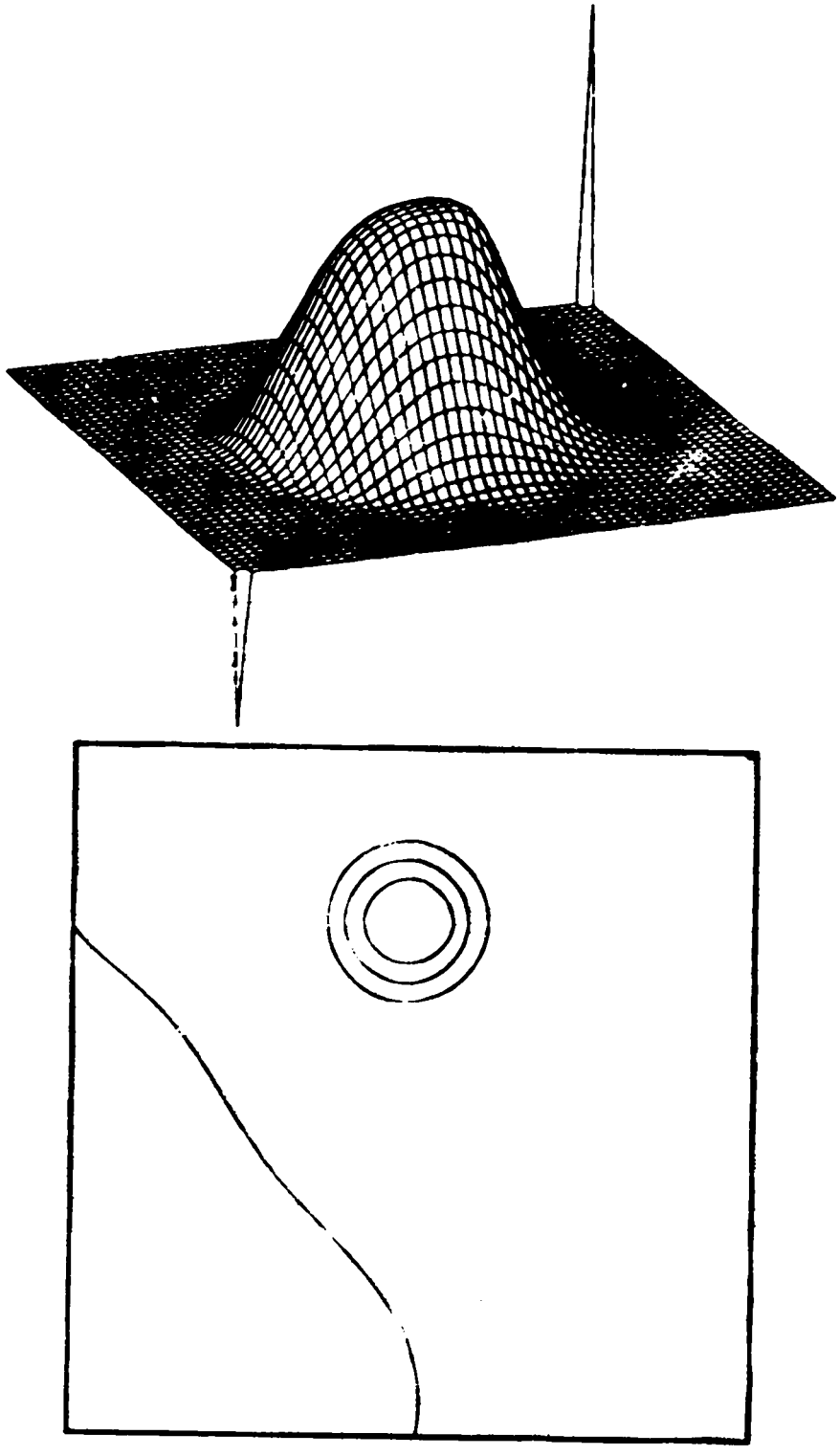


Figure F.28: The Runge-Kutta HOG method with the van Leer limiter solution for the rotating cone shows the better resolution of this limiter.

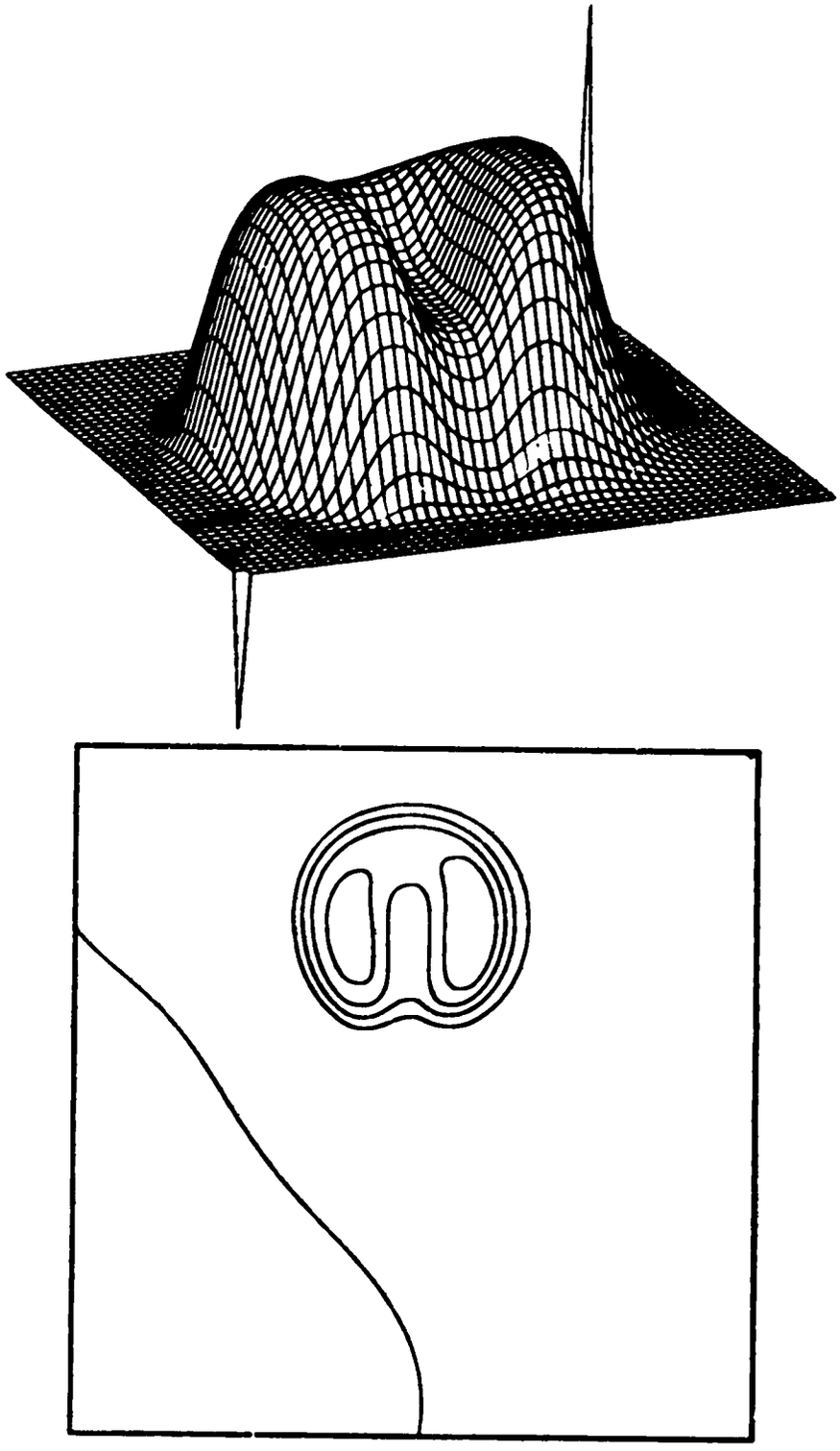


Figure F.29: The Runge-Kutta HOC method with the van Leer limiter solution for the rotating slotted cylinder shows the better resolution of this limiter.

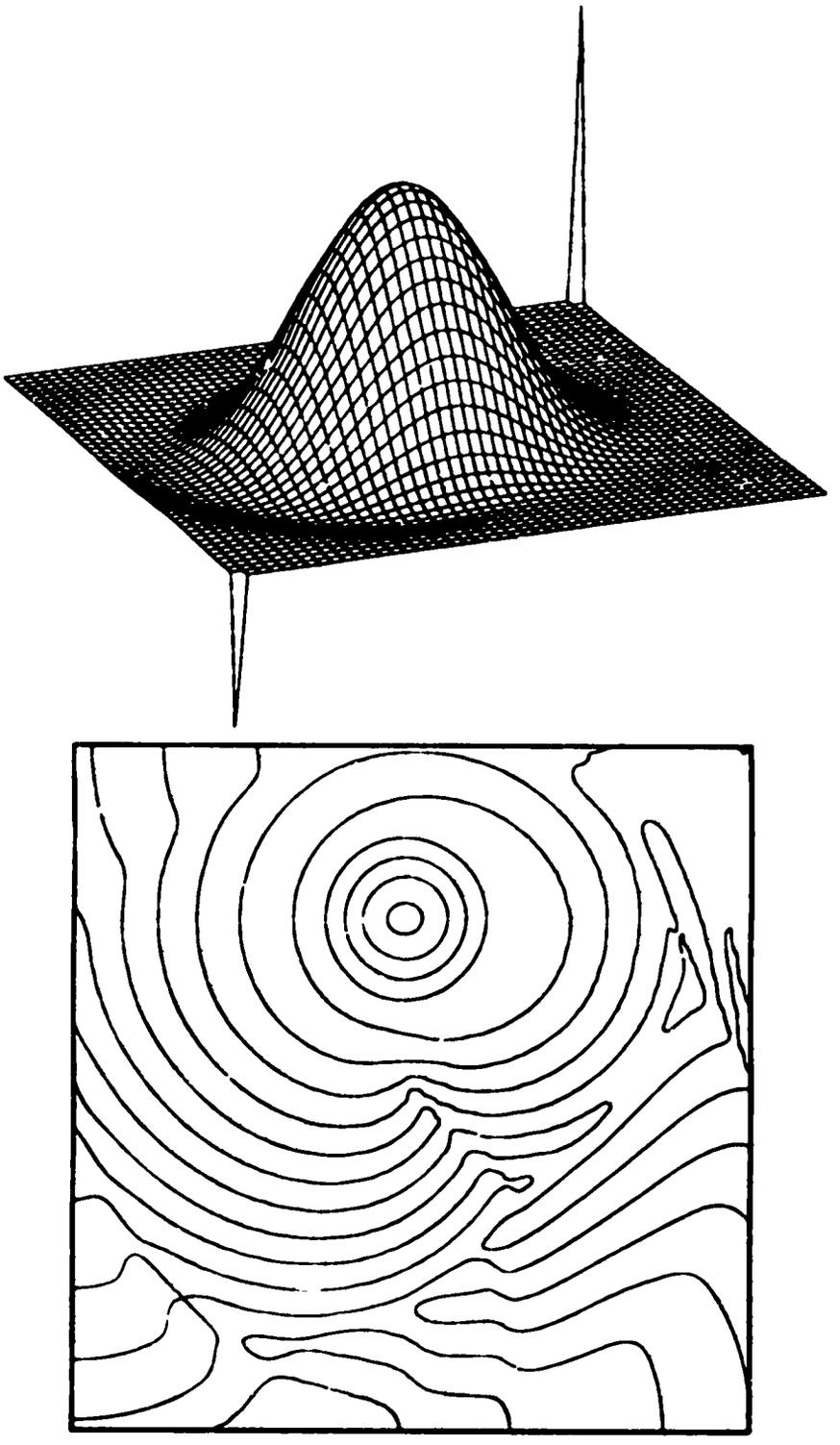


Figure F.30: The Runge-Kutta HOC method with the generalized average limiter $n = 2$ solution for the rotating cone shows the better resolution of this limiter, but the non-monotonic behavior.

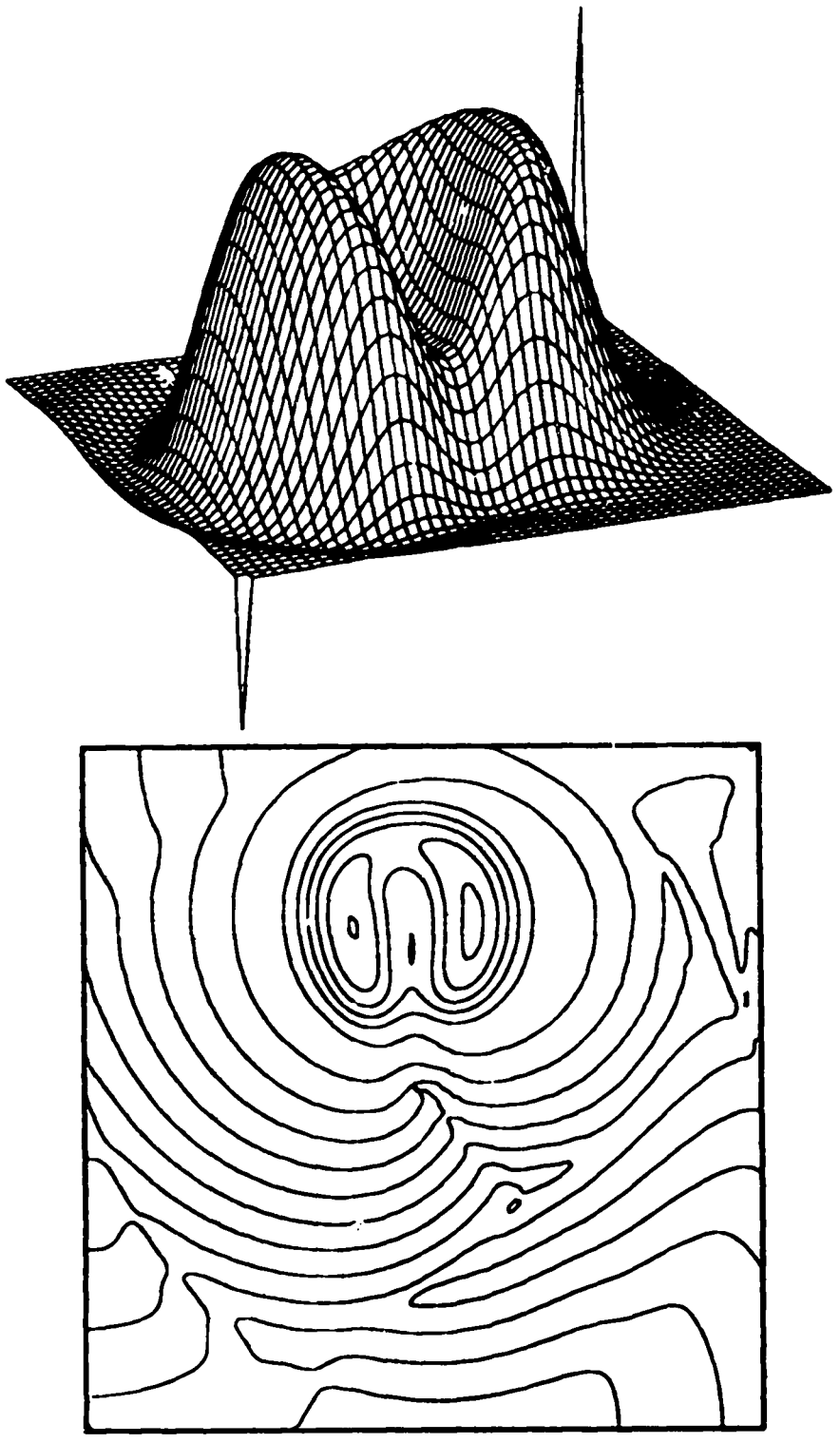


Figure F.31: The Runge-Kutta HOG method with the generalized average limiter $n = 2$ solution for the rotating slotted cylinder shows the better resolution of this limiter, but the non-monotonic behavior.

References

[1] S. V. Patankar. *Numerical Heat Transfer and Fluid Flow.* Hemisphere, 1980.

[2] D. R. Liles and W. H. Reed. A semi-implicit method for two-phase fluid dynamics. *Journal of Computational Physics*, 26:390-407, 1978.

[3] F. H. Harlow and A. A. Amsden. A numerical fluid dynamics calculation method for all flow speeds. *Journal of Computational Physics*, 8:197-213, 1971.

[4] E. S. Oran and J. P. Boris. *Numerical Simulation of Reactive Flow.* Elsiver, 1987.

[5] W. J. Rider. A comparison of TVD Lax-Wendroff schemes. Technical Report LA-UR-91-2770, Los Alamos National Laboratory, 1992. Accepted in Communications in Applied Numerical Methods.

[6] W. J. Rider and D. R. Liles. A generalized flux-corrected transport algorithm. I: A finite difference formulation. Technical Report LA-UR-90-3725, Los Alamos National Laboratory, 1990. submitted to the Journal of Computational Physics.

[7] W. J. Rider. The use of approximate Riemann solvers with Godunov's method in Lagrangian coordinates. Technical Report LA-UR-91-2555, Los Alamos National Laboratory, 1991. submitted to Computers and Fluids.

[8] W. J. Rider. A generalized flux-corrected transport algorithm. II: A geometric approach. Technical Report LA-UR-91-2769, Los Alamos National Laboratory, 1991. submitted to the SIAM Journal on Scientific and Statistical Computing.

[9] W. J. Rider. Limiters in the high resolution solution of hyperbolic conservation laws. Technical Report LA-UR-91-3568, Los Alamos National Laboratory, 1991. submitted to SIAM Journal on Numerical Analysis.

[10] W. J. Rider. Methods for extending high resolution schemes to systems of hyperbolic conservation laws. Technical Report LA-UR-91-3286, Los Alamos National Laboratory, 1991. submitted to International Journal of Numerical Methods in Engineering.

[11] W. J. Rider. Cell averages or point values ? On reconstruction methods for the solution of hyperbolic conservation laws. Technical Report LA-UR-91-3567, Los Alamos National Laboratory, 1991. Submitted to Journal of Computational Physics.

[12] J. J. Monaghan. Why particle methods work. *SIAM Journal on Scientific and Statistical Computing*, 3:422-433, 1982.

[13] A. J. Majda. Vorticity, turbulence and acoustics in fluid flow. *SIAM Review*, 33:349-388, 1991.

[14] J. Glimm. Nonlinear and stochastic phenomena: The grand challenge for partial differential equations. *SIAM Review*, 33:626-643, 1991.

[15] M. Brio and C. C. Wu. An upwind differencing scheme for the equations of ideal magnetohydrodynamics. *Journal of Computational Physics*, 75:400–422, 1988.

[16] S. Osher. IBM Third Computational Fluid Dynamics Short Course, Monterey CA, May 1991.

[17] P. D. Lax. Shock waves and entropy. In E. H. Zarantonello, editor, *Contributions to Nonlinear Functional Analysis*, pages 603–634. Academic Press, 1971.

[18] P. D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*. SIAM, 1972.

[19] J. Smoller. *Shock Waves and Reaction-Diffusion Equations*. Springer-Verlag, 1982.

[20] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics*. Pergamon Press, 1987.

[21] D. Mihalas and B. W. Mihalas. *Foundations of Radiation Hydrodynamics*. Oxford, 1984.

[22] J. J. Duderstadt and W. R. Martin. *Transport Theory*. Wiley-Interscience, 1979.

[23] A. J. Chorin and G. Marsden. *A Mathematical Introduction to Fluid Mechanics*. Springer-Verlag, 1990.

[24] Jr. J. D. Anderson. *Modern Compressible Flow*. McGraw Hill Book Company, 1982.

[25] R. Courant and K. O. Friedrichs. *Supersonic Flow and Shock Waves*. Interscience, 1948.

[26] A. J. Chorin. Numerical solution of Boltzmann's equation. *Communications on Pure and Applied Mathematics*, 25:171–186, 1972.

[27] P. R. Woodward. Piecewise-parabolic methods for astrophysical fluid dynamics. In K.-H. A Winkler and M. L. Norman, editors, *Astrophysical Radiation Hydrodynamics*, pages 245–326, 1986.

[28] D. D. Joseph and L. Preziosi. Heat waves. *Reviews in Modern Physics*, 61:41–73, 1989.

[29] H. D. Weymann. Finite speed of propagation in heat conduction, diffusion, and viscous shear motion. *American Journal of Physics*, 35:488–496, 1967.

[30] A. Harten, P. D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Review*, 25:35–61, 1983.

[31] R. D. Richtmyer and K. W. Morton. *Difference Methods for Initial Value Problems*. Wiley-Interscience, 1967.

[32] P. J. Roache. *Computational Fluid Dynamics*. Hermosa, 1976.

[33] D. Potter. *Computational Physics*. Wiley-Interscience, 1973.

[34] D. A. Anderson, J. C. Tannehill, and R. H. Pletcher. *Computational Fluid Mechanics and Heat Transfer*. Hemisphere, 1984.

[35] C. Hirsch. *Numerical Computation of Internal and External Flows: Volume 1*. Wiley-Interscience, 1988.

[36] C. Hirsch. *Numerical Computation of Internal and External Flows: Volume 2*. Wiley-Interscience, 1988.

[37] C. A. J. Fletcher. *Computational Techniques for Fluid Dynamics: Volume I*. Springer-Verlag, 1988.

[38] C. A. J. Fletcher. *Computational Techniques for Fluid Dynamics: Volume II*. Springer-Verlag, 1988.

[39] G. A. Sod. *Numerical Methods in Fluid Dynamics*. Cambridge, 1985.

[40] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser, 1990.

[41] G. Sod. A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws. *Journal of Computational Physics*, 27:1–31, 1978.

[42] J. P. Boris. New directions in computational fluid dynamics. *Annual Reviews in Fluid Mechanics*, 21:345–385, 1989.

[43] A. Rizzi and B. Engquist. Selected topics in the theory and practice of computational fluid dynamics. *Journal of Computational Physics*, 72:1–69, 1987.

[44] P. Woodward and P. Colella. The numerical simulation of two-dimensional fluid flow with strong shocks. *Journal of Computational Physics*, 54:115–173, 1984.

[45] H. C. Yee. Upwind and symmetric shock-capturing schemes. Technical Report NASA TM-89464, NASA, 1987.

[46] R. B. Rood. Numerical advection algorithms and their role in atmospheric transport and chemistry models. *Reviews of Geophysics*, 25:71–100, 1987.

[47] S. Osher and P. K. Sweby. Recent developments in the numerical solution of nonlinear conservation laws. In A Iserles and M. J. D. Powell, editors, *The State of the Art in Numerical Analysis*, pages 682–701, 1987.

[48] H. Lomax. CFD in the 1980's from one point of view. In D. Kwak, editor, *Proceedings of the AIAA Tenth Computational Fluid Dynamics Conference*, pages 1–9, 1991. AIAA Paper 91–1526.

[49] A. J. Przekwas and H. Q. Yang. Advanced CFD methodology for fast transients encountered in nonlinear combustion instability problems: SBIR phase I final report. Technical Report Report 4065/1, CFD Research Corporation, 1989.

[50] J. D. Baum and J. N. Levine. A critical study of numerical methods for the solution of nonlinear hyperbolic equations for resonance systems. *Journal of Computational Physics*, 58:1-28, 1985.

[51] L. F. Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with applications to stresses in a masonry dam. *Transactions of the Royal Society of London*, 210:307-357, 1910.

[52] R. Courant, K. Friedrichs, and H. Lewy. On the partial differential equations of mathematical physics. *Mathematische Annalen*, 100:32-74, 1928.

[53] J. VonNeumann and R. D. Richtmyer. A method for the numerical calculation of hydrodynamic shocks. *Journal of Applied Physics*, 21:232-237, 1950.

[54] R. Courant, E. Issacson, and M. Rees. On the solution of nonlinear hyperbolic differential equations by finite differences. *Communications on Pure and Applied Mathematics*, 5:243-255, 1952.

[55] P. D. Lax. Weak solutions of nonlinear hyperbolic equations and their numerical computation. *Communications on Pure and Applied Mathematics*, 7:159-193, 1954.

[56] S. K. Godunov. Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics. *Matematicheski Sbornik*, 47:271-306, 1959.

[57] S. K. Godunov, A. V. Zabrodyn, and G. P. Prokopov. A computational scheme for two-dimensional nonstationary problems of gas dynamics and calculation of the flow from a shock wave approaching steady-state. *USSR Journal of Computational Mathematics and Mathematical Physics*, 1:1187-1219, 1961.

[58] P. D. Lax and B. Wendroff. Systems of conservation laws. *Communications on Pure and Applied Mathematics*, 13:217-237, 1960.

[59] J. P. Boris and D. L. Book. Flux-corrected transport I. SHASTA, a fluid transport algorithm that works. *Journal of Computational Physics*, 11:38-69, 1973.

[60] B. van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. *Journal of Computational Physics*, 32:101-136, 1979.

[61] A. Harten. On a class of high resolution total-variation-stable finite-difference schemes. *SIAM Journal on Numerical Analysis*, 21:1-23, 1984.

[62] S. T. Zalesak. Fully multidimensional flux-corrected transport algorithms for fluids. *Journal of Computational Physics*, 31:335-362, .979.

[63] P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43:357-372, 1981.

[64] A. Harten, B. Engquist, S. Osher, and S. Chakravarthy. Uniformly high order accurate essentially non-oscillatory schemes, III. *Journal of Computational Physics*, 71:231-303, 1987.

[65] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77:439-471, 1988.

[66] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes II. *Journal of Computational Physics*, 83:32-78, 1989.

[67] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge, 1986.

[68] R. Menikoff and B. J. Plohr. The Riemann problem for fluid flow of real materials. *Reviews in Modern Physics*, 61:75-129, 1989.

[69] P. D. Lax. Nonlinear partial differential equations and computing. *SIAM Review*, 11:7-19, 1969.

[70] J. Glimm. The interaction of nonlinear hyperbolic waves *Communications on Pure and Applied Mathematics*, 41:569-590, 1988.

[71] H. B. Stewart and B. Wendroff. Two-phase flow: Models and methods. *Journal of Computational Physics*, 56:363-409, 1984.

[72] V. H. Ransom and D. L. Hicks. Hyperbolic two-pressure models in two-phase flow. *Journal of Computational Physics*, 53:124-151, 1984.

[73] A. Harten, J. M. Hyman, and P. D. Lax. On finite difference approximations and entropy conditions for shocks. *Communications on Pure and Applied Mathematics*, 29:297-322, 1976.

[74] M. Vinokur. An analysis of finite-difference and finite-volume formulations of conservation laws. *Journal of Computational Physics*, 81:1-52, 1989.

[75] I. I. Glass. Some aspects of shock-wave research. *AIAA Journal*, 25:214-229, 1987.

[76] W. C. Reynolds. The potential and limitations of direct and large eddy simulations. *Lecture Notes in Physics*, 357:313-343, 1990.

[77] J. P. Boris. On large eddy simulation using subgrid turbulence models. *Lecture Notes in Physics*, 357:344-351, 1990.

[78] D. H. Porter, P. R. Woodward, W. Yang, and Q. Mei. Simulation and visualization of compressible convection in two and three dimensions. In J. R. Buchler and S. T. Gottesmann, editors, *Nonlinear Astrophysical Fluid Dynamics*, pages 234-258, 1990.

[79] D. H. Porter, A. Pouquet, and P. R. Woodward. A numerical study of supersonic homogeneous turbulence, 1992. preprint.

[80] S. V. Patankar. Recent developments in computational heat transfer. *Journal of Heat Transfer*, 110:1037-1045, 1988.

[81] B. P. Leonard. Simple high-accuracy resolution program for convective modelling of discontinuities. *International Journal for Numerical Methods in Fluids*, 8:1291-1318, 1988.

[82] B. P. Leonard and S. Mokhtari. Beyond first-order upwinding the ULTRA-SHARP alternative for non-oscillatory steady-state simulation of convection. *International Journal for Numerical Methods in Engineering*, 30:729-766, 1990.

[83] B. P. Leonard and H. S. Niknafs. Sharp monotonic resolution of discontinuities without clipping of narrow extrema. *Computers and Fluids*, 19:141-154, 1991.

[84] B. P. Leonard. The ULTIMATE convective difference scheme applied to unsteady one-dimensional advection. *Computer Methods in Applied Mechanics and Engineering*, 88:17-74, 1991.

[85] J. B. Bell, P. Colella, and H. M. Glaz. A second-order projection method of the incompressible Navier-Stokes equations. *Journal of Computational Physics*, 85:257-283, 1989.

[86] J. B. Bell, L. Howell, and P. Colella. An efficient second-order projection method for viscous incompressible flow. In D. Kwak, editor, *Proceedings of the AIAA Tenth Computational Fluid Dynamics Conference*, pages 360-367, 1991. AIAA Paper 91-1560.

[87] D. B. Kothe, R. C. Mjolsness, and M. D. Torrey. RIPPLE: A computer program of incompressible flows with free surfaces. Technical Report LA-1200-7-MS, Los Alamos National Laboratory, 1991.

[88] J. F. Hawley, L. L. Smarr, and J. R. Wilson. A numerical study of nonspherical black hole accretion. II. a finite differencing and code calibration. *Astrophysical Journal Supplementary Series*, 55:211-246, 1984.

[89] B. Fryxell, E. Müller, and D. Arnett. Instabilities and clumping in SN 1987a. I. Early evolution in two dimensions. *Astrophysical Journal*, 367:619-634, 1991.

[90] E. Müller, W. Hildebrant, M. Orio, P. Höflich, R. Mönchmeyer, and B. Fryxell. Mixing and fragmentation in supernova envelopes. *Astronomy and Astrophysics*, 220:167-176, 1989.

[91] E. Müller, B. Fryxell, and D. Arnett. Instability and clumping in SN 1987A. *Astronomy and Astrophysics*, 251:505-514, 1991.

[92] D. Arnett, B. Fryxell, and E. Müller. Instability and nonradial motion in SN 1987A. *Astrophysical Journal*, 341:L63-L66, 1989.

[93] R. L. Carpenter, K. K. Droegemeier, P. R. Woodward, and C. E. Hane. Application of the piecewise parabolic method (PPM) to meteorological modeling. *Monthly Weather Review*, 118:586-612, 1990.

[94] D. J. Allen, A. R. Douglass, R. B. Rood, and P. L. [illegible]e. Application of a monotonic upstream-biased transport scheme to three-dimensional constituent transport calculations. *Monthly Weather Review*, 119:2456-2464, 1991.

[95] M. J. Fritts, W. P. Crowley, and H. Trease, editors. *The Free Lagrange Method*, volume 238 of *Lecture Notes in Physis*. Springer-Verlag, 1985.

[96] H. Q. Yang. Characteristics-based high-order accurate and nonoscillatory numerical method for hyperbolic heat conduction. *Numerical Heat Transfer, Part B*, 18:221-241, 1990.

[97] V. Shankar, W. F. Hall, and A. H. Mohammadian. A time-domain differential solver for electromagnetic scattering problems. *Proceedings of the IEEE*, 77:709-721, 1989.

[98] V. Shankar, W. F. Hall, A. H. Mohammadian, and S. Chakravarthy. Applications of computation fluid dynamics-based methods to problems in computational science. *Computer Systems in Engineering*, 1:7-22, 1991.

[99] I. Toumi and P. Raymond. Numerical method for two-phase flow discontinuity propagation calculation. In M. L. Hall, editor, *Advances in Nuclear Engineering Computation and Radiation Shielding*, page 53, 1989.

[100] C. P. Tzanos. Central difference-like approximations for the solution of the convection-diffusion equation. *Numerical Heat Transfer, Part B*, 17:97-112, 1990.

[101] W. J. Rider and S. B. Woodruff. High-order solute tracking in two-phase thermal hydraulics. In H. A. Dwyer, editor, *Proceedings of the Fourth International Symposium on Computational Fluid Dynamics*, pages 957-962, 1991.

[102] J. A. Trangenstein and P. Colella. A higher-order Godunov method for modeling finite deformation in elastic-plastic solids. *Communications on Pure and Applied Mathematics*, 44:41-100, 1991.

[103] J. B. Bell, G. R. Shubin, and J. A. Trangenstein. A method for reducing numerical dispersion in two-phase black-oil reservoir simulation. *Journal of Computational Physics*, 65:71-106, 1986.

[104] J. B. Bell, P. Colella, and J. A. Trangenstein. High order Godunov methods for general systems of hyperbolic conservation laws. *Journal of Computational Physics*, 82:362-397, 1989.

[105] Y. Brenier and J. Jaffé. Upstream differencing for multiphase flow in reservoir simulation. *SIAM Journal on Numerical Analysis*, 28:685-696, 1991.

[106] A. Jameson. A nonoscillatory shock capturing scheme using flux limited dissipation. In B. Engquist et al., editor, *Lectures in Applied Mathematics*, pages 48-65, 1985.

[107] J. R. Buchler and P. Whalen. Experiments with artificial viscosity. In J. R. Buchler, editor, *The Numerical Modelling of Nonlinear Stellar Pulsations*, pages 269-288, 1990.

[108] T. Y. Hou and P. D. Lax. Dispersive approximations in fluid dynamics. *Communications on Pure and Applied Mathematics*, 44:1-40, 1991.

[109] E. M. Murman and J. D. Cole. Calculation of plane steady transonic flows. *AIAA Journal*, 9:114-121, 1971.

[110] H. C. Yee, R. F. Warming, and A. Harten. Implicit total variation diminishing (TVD) schemes for steady-state applications. *Journal of Computational Physics*, 57:327-360, 1985.

[111] P. D. Lax and B. Wendroff. Difference schemes for hyperbolic equations with high order of accuracy. *Communications on Pure and Applied Mathematics*, 17:381-398, 1964.

[112] A. Harten. From artificial viscosity to ENO schemes. In J. R. Buchler, editor, *The Numerical Modelling of Nonlinear Stellar Pulsations*, pages 239-262, 1990.

[113] R. D. Richtmyer. A survey of difference methods for non-steady gas dynamics. Technical Report NCAR Tech Note 63-2, NCAR, 1963.

[114] S. Z. Burstein. Finite-difference calculations for hydrodynamic flows containing discontinuities. *Journal of Computational Physics*, 1:198-222, 1966.

[115] R. W. MacCormack. The effect of viscosity in hypervelocity impact cratering, 1969. AIAA Paper 69-354.

[116] R. M. Beam and R. F. Warming. An implicit factored scheme for the compressible Navier-Stokes equations. *AIAA Journal*, 16:393-402, 1978.

[117] M. J. Berger and P. Colella. Local adaptive mesh refinement for shock hydrodynamics. *Journal of Computational Physics*, 82:64-84, 1989.

[118] B. van Leer. *A Choice of Difference Schemes for Ideal Compressible Flow.* PhD thesis, University of Leiden, 1970.

[119] B. van Leer. Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme. *Journal of Computational Physics*, 14:361-370, 1974.

[120] B. van Leer. Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. *Journal of Computational Physics*, 23:276-299, 1977.

[121] P. Colella. Glimm's method for gas dynamics. *SIAM Journal on Scientific and Statistical Computing*, 3:76-110, 1982.

[122] P. Colella and P. Woodward. The piecewise parabolic method (PPM) for gas-dynamical simulations. *Journal of Computational Physics*, 54:174-201, 1984.

[123] P. Colella. A direct Eulerian MUSCL scheme for gas dynamics. *SIAM Journal on Scientific and Statistical Computing*, 6:104-117, 1985.

[124] P. Colella and H. M. Glaz. Efficient solution algorithms for the Riemann problem for real gases. *Journal of Computational Physics*, 58:264–289, 1985.

[125] J. L. Steger and R. F. Warming. Flux vector splitting of the inviscid gas-dynamics equations with application to finite-difference methods. *Journal of Computational Physics*, 40:263–293, 1981.

[126] B. van Leer. Flux-vector splitting for the Euler equations. *Lecture Notes in Physics*, 170:507–512, 1981.

[127] B. Engquist and S. Osher. One-sided difference approximations for nonlinear conservation laws. *Mathematics of Computation*, 36:321–351, 1981.

[128] B. Einfeldt. On Godunov-type methods for gas dynamics. *SIAM Journal on Numerical Analysis*, 25:294–318, 1988.

[129] A. Bourlioux, A. J. Majda, and V. Roytburd. Theoretical and numerical structure for unstable one-dimensional detonations. *SIAM Journal of Applied Mathematics*, 51:303–343, 1991.

[130] A. Harten. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 49:357–393, 1983.

[131] P. L. Roe. Generalized formulation of TVD Lax-Wendroff schemes. Technical Report NASA CR-172478/ICASE Report 84-53, NASA, 1984.

[132] P. K. Sweby. High-resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM Journal on Numerical Analysis*, 21:995–1011, 1984.

[133] S. F. Davis. A simplified TVD finite difference scheme via artificial viscosity. *SIAM Journal on Scientific and Statistical Computing*, 8:1–18, 1987.

[134] H. C. Yee. Construction of explicit and implicit symmetric TVD schemes and their applications. *Journal of Computational Physics*, 68:151–179, 1987.

[135] J. B. Goodman and R. J. Leveque. A geometric approach to high resolution TVD schemes. *SIAM Journal on Numerical Analysis*, 25:268–284, 1988.

[136] A. Harten and S. Osher. Uniformly high-order accurate nonoscillatory schemes. I. *SIAM Journal on Numerical Analysis*, 24:279–309, 1987.

[137] A. Harten. ENO schemes with subcell resolution. *Journal of Computational Physics*, 83:148–184, 1989.

[138] A. G. Godfrey, C. R. Mitchell, and R. W. Walters. Practical aspects of spatially high accurate methods. In *30th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 1992. AIAA-92-0054.

[139] L. J. Durlofsky, B. Engquist, and S. Osher. Triangle based adaptive stencils for the solution of hyperbolic conservation laws, 1991. to be published

[140] J. P. Boris, D. L. Book, and K Hain. Flux-corrected transport II: Generalizations of the method. *Journal of Computational Physics*, 18:248–283, 1975.

[141] J. P. Boris and D. L. Book. Flux-corrected transport III. minimal-error FCT algorithms. *Journal of Computational Physics*, 20:397–431, 1976.

[142] J. P. Boris and D. L. Book. *Solution of Continuity Equations by the Method of Flux-Corrected Transport*, volume 16, pages 85–129. Academic Press, 1976.

[143] S. T. Zalesak. High order "ZIP" differencing of convective terms. *Journal of Computational Physics*, 40:497–508, 1981.

[144] R. Löhner, K. Morgan, J. Peraire, and M. Vahdati. Finite element flux-corrected transport (FEM-FCT) for the Euler and Navier-Stokes equations. *International Journal for Numerical Methods in Fluids*, 7:1093–1109, 1987.

[145] H. C. Yee. A class of high-resolution explicit and implicit shock-capturing methods. Technical Report NASA TM-101088, NASA, 1989.

[146] A. Harten and G. Zwas. Self-adjusting hybrid schemes for shock computation. *Journal of Computational Physics*, 6:568–583, 1972.

[147] B. van Leer. Upwind-difference methods for aerodynamic problems governed by the Euler equations. In B. Engquist et al., editor, *Lectures in Applied Mathematics*, volume 22, pages 327–336, 1985.

[148] D. L. Williamson and P. J. Rasch. Two-dimensional semi-Lagrangian transport with shape-preserving interpolation. *Monthly Weather Review*, 117:102–129, 1989.

[149] P. J. Rasch and D. L. Williamson. On shape-preserving interpolation and semi-Lagrangian transport. *SIAM Journal on Scientific and Statistical Computing*, 11:656–687, 1990.

[150] J. B. Goodman and R. J. LeVeque. On the accuracy of stable schemes for 2d scalar conservation laws. *Mathematics of Computation*, 45:15–21, 1985.

[151] P. L. Roe. A basis for upwind differencing of the two-dimensional unsteady euler equations. In K. W. Morton and M. J. Baines, editors, *Numerical Methods for Fluid Dynamics II*, pages 55–83, 1986.

[152] R. J. DiPerna. Convergence of approximate solutions to conservation laws. *Archives for Rational Mechanics and Analysis*, 82:27–70, 1983.

[153] R. J. DiPerna. Oscillations in solutions to nonlinear differential equations. In C. Dafermos, J. L. Erickson, D. Kinderlehrer, and M. Slemrod, editors, *The IMA Volumes in Mathematics and its Applications*, volume 2, pages 23–34. Springer-Verlag, 1986.

[154] H. C. Yee, P. K. Sweby and D. F. Griffiths. Dynamical approach study of spurious steady state numerical solutions of nonlinear differential equations I the dynamics of time discretization and its implications for algorithm development in computational fluid dynamics. *Journal of Computational Physics*, 97:249–310, 1991.

[155] S. Osher. Riemann solvers, the entropy condition, and difference approximations. *SIAM Journal on Numerical Analysis*, 21:217–235, 1984.

[156] N. N. Yanenko. *The Method of Fractional Steps*. Springer Verlag, 1971.

[157] C. W. Hirt, A. A. Amsden, and J. L. Cook. An arbitrary Lagrangian Eulerian computing method for all flow speeds. *Journal of Computational Physics*, 14:227–253, 1974.

[158] B. van Leer. On the relation between the upwind-differencing schemes of Godunov, Engquist-Osher and Roe. *SIAM Journal on Scientific and Statistical Computing*, 5:1–20, 1984.

[159] G. D. van Albada, B. van Leer, and W. W. Roberts. A comparative study of computational methods in cosmic gas dynamics. *Astronomy and Astrophysics*, 108:76–84, 1982.

[160] C. W. Shu. Total-variation-diminishing time discretizations. *SIAM Journal on Scientific and Statistical Computing*, 9:1073–1084, 1988.

[161] A. Harten. Preliminary results on the extension of ENO schemes to two-dimensional problems. In *Nonlinear Hyperbolic Problems*, pages 23–40, 1986.

[162] A. Harten. On high-order accurate interpolation for non-oscillatory shock capturing schemes. Technical Report MRC Tech. Rep 2829, University of Wisconsin, 1985.

[163] F. B. Hildebrand. *Introduction to Mathematical Analysis*. Dover, 1974.

[164] B. Engquist, P. Lötstedt, and B. Sjögreen. Nonlinear filters for efficient shock computation. *Mathematics of Computation*, 52:509–537, 1989.

[165] F. Lafon and S. Osher. High order filtering methods for approximating hyperbolic systems of conservation laws. *Journal of Computational Physics*, 96:110–142, 1991.

[166] H. C. Yee, G. H. Klopfer, and J.-L. Montagne. High-resolution shock-capturing schemes for inviscid and viscous hypersonic flows. *Journal of Computational Physics*, 88:31–61, 1990.

[167] J. Y. Yang. Uniformly second-order accurate essentially nonoscillatory scheins for the Euler equations. *AIAA Journal*, 28:2069–2076, 1990.

[168] A. Jameson and P. D. Lax. Conditions for the construction of multi-point total variation diminishing difference schemes. *Applied Numerical Mathematics*, 2:335–345, 1986.

[169] C.-W. Shu. TVB uniformly high-order schemes for conservation laws. *Mathematics of Computation*, 49:105-121, 1987.

[170] T. Ikeda and T. Nakagawa. On the SHASTA FCT algorithm for the equation $\partial\rho/\partial t + \partial(v(\rho))/\partial x = 0$. *Mathematics of Computation*, 33:1157-1169, 1979.

[171] C. R. DeVore. Flux-corrected transport techniques for multidimensional compressible magnetohydrodynamics. *Journal of Computational Physics*, 92:142-160, 1991.

[172] E. S. Oran, J. P. Bois, and D. A. Jones. Reactive-flow computations on a connection machine. In K. W. Morton, editor, *Twelfth International Conference on Numerical Methods in Fluid Dynamics*, pages 318-322, 1990.

[173] B. E. McDonald. Flux-corrected pseudospectral method for scalar hyperbolic conservation laws. *Journal of Computational Physics*, 82:413-428, 1989.

[174] P. Steinle and R. Morrow. An implicit flux-corrected transport algorithm. *Journal of Computational Physics*, 80:61-71, 1989.

[175] D. L. Book, editor. *Finite-Difference Techniques for Vectorized Fluid Dynamic Calculations*. Springer-Verlag, 1981.

[176] P. L. Roe. Some contributions to the modelling of discontinuous flows. In B. Engquist et al., editor, *Lectures in Applied Mathematics*, volume 22, pages 163-193, 1985.

[177] J. J. Gottlieb and C. P. T. Groth. Assessment of Riemann solvers for unsteady one-dimensional inviscid flows of perfect gases. *Journal of Computational Physics*, 78:437-458, 1988.

[178] A. Harten. Short course: Numerical methods for hyperbolic conservation laws, 1986. ICASE Internal Report, Document Number 33.

[179] S. Osher. Convergence of generalized MUSCL schemes. *SIAM Journal on Numerical Analysis*, 22:947-961, 1985.

[180] S. Osher and S. Chakravarthy. High resolution schemes and the entropy condition. *SIAM Journal on Numerical Analysis*, 21:955-983, 1984.

[181] C.-D. Munz. On the numerical dissipation of high resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 77:18-39, 1988.

[182] A. Harten and J. M. Hyman. A self-adjusting grid for the computation of weak solutions of hyperbolic conservation laws. *Journal of Computational Physics*, 50:235-269, 1983.

[183] A. Harten. The artificial compression method for computation of shocks and contact discontinuities: III. Self-adjusting hybrid schemes. *Mathematics of Computation*, 32:363-389, 1978.

[184] B. Cockburn. Quasimonotone schemes for scalar conservation laws. Part I. *SIAM Journal on Numerical Analysis*, 26:1325–1341, 1989.

[185] J. P. Vila. High-order schemes and entropy condition for nonlinear hyperbolic systems of conservation laws. *Mathematics of Computation*, 50:53–73, 1988.

[186] P. K. Sweby. Flux limiters. In F. Angrand, A. Dervieux, J. A. Desideri, and R. Glowinski, editors, *Numerical Methods for the Euler Equations of Fluid Dynamics*, pages 48–65, 1985.

[187] P. K. Sweby. High resolution TVD schemes using flux limiters. In B. Engquist et al., editor, *Lectures in Applied Mathematics*, volume 22, pages 289–309, 1985.

[188] S. Osher and C.-W. Shu. High-order essentially nonoscillatory schemes for Hamilton-Jacobi equations. *SIAM Journal on Numerical Analysis*, 28:907–922, 1991.

[189] S. F. Davis. Simplified second-order Godunov-type methods. *SIAM Journal on Scientific and Statistical Computing*, 9:445–473, 1988.

[190] Z. Wang and B. E. Richards. High resolution schemes for steady flow computation. *Journal of Computational Physics*, 97:53–72, 1991.

[191] A. Suresh and H. T. Huynh. Numerical experiments on a new class of nonoscillatory schemes. In *30th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 1992. AIAA-92-0421.

[192] H. Yang. An artificial compression method for ENO schemes: The slope modification method. *Journal of Computational Physics*, 89:125–160, 1990.

[193] D.-K. Mao. A treatment of discontinuities in shock-capturing finite difference methods. *Journal of Computational Physics*, 92:422–455, 1991.

[194] E. Tadmor. Convenient total variation diminishing conditions for nonlinear difference schemes. *SIAM Journal on Numerical Analysis*, 25.1002–1014, 1988.

[195] W. A. Mulder. A note of the use of symmetric Gauss-Seidel for the steady upwind differenced Euler equations. *SIAM Journal on Scientific and Statistical Computing*, 11:389–397, 1990.

[196] W. A. Mulder and B. van Leer. Experiments with implicit upwind methods for the Euler equations. *Journal of Computational Physics*, 59:232–248, 1985.

[197] [illegible]-C. Chieng. Characteristic-based flux limiters of an essentially third-order flux-splitting method for hyperbolic conservation laws. *International Journal for Numerical Methods in Fluids*, 13:287–307, 1991.

[198] W. A. Mulder. Multigrid relaxation for the Euler equations. *Journal of Computational Physics*, 60:235–252, 1985.

[199] A. Harten and S. R. Chakravarthy. Multi-dimensional ENO schemes for general geometries. Technical Report CAM Report 91-16, UCLA, 1991.

[200] H. Nessyahu and E. Tadmor. Non-oscillatory central differencing for hyperbolic conservation laws. *Journal of Computational Physics*, 87:408-463, 1990.

[201] D. Dubois and H. Prade. *Fuzzy Sets and Systems*. Academic Press. 1980.

[202] R. L.-P. Chang and T. Pavlidis. Applications of fuzzy sets in curve fitting. *Fuzzy Sets and Systems*, 2:67-74, 1988.

[203] C.-W. Shu, T. A. Zang, G. Erlebacher, D. Whitaker, and S. Osher. High-order ENO schemes applied to two- and three-dimensional compressible flow. *Applied Numerical Mathematics*, 9:45-71, 1992.

[204] W. Benz. Applications of smooth particle hydrodynamics (SPH) to astrophysical problems. *Computer Physics Communications*, 48:97-105, 1988.

[205] W. Benz. Smooth particle hydrodynamics: A review. In J. R. Buchler, editor, *The Numerical Modelling of Nonlinear Stellar Pulsations*, pages 269-288, 1990.

[206] J. J. Monaghan. An introduction to SPH. *Computer Physics Communications*, 48:89-96, 1988.

[207] W. A. Mulder. Computation of quasi-steady gas flow in a spiral galaxy by means of a multigrid method. *Astronomy and Astrophysics*, 156:354-380, 1986.

[208] A. Jameson and W. Schmidt. Some recent developments in numerical methods for transonic flows. *Computer Methods in Applied Mechanics and Engineering*, 51:467-493, 1985.

[209] A. Jameson. Computational transonics. *Communications on Pure and Applied Mathematics*, 41:507-549, 1988.

[210] B. Engquist and Q. Q. Huynh. Iterative gradient-Newton type methods for steady shock computations. In S. Gomez, J. P. Hennart, and R. A. Tapia, editors, *Advances in Numerical Partial Differential Equations and Optimization*, pages 60 75. SIAM, 1991.

[211] T. W. Roberts. The behavior of flux difference splitting schemes near slowly moving shock waves. *Journal of Computational Physics*, 90:141-160, 1990.

[212] D. Rotman. Shock wave effects on a turbulent flow. *Phys. Fl. A*, 3:1792-1806, 1991.

[213] P. R. Woodward. Picewise-parabolic methods for systems of hyperbolic conservation laws in multiprocessing environments. Technical Report DOE/ER/25035-2, University of Minnesota, 1990.

[214] I.-L. Chern and P. Colella. A conservative front tracking method for hyperbolic conservation laws. Technical Report UCRL-97200, Lawrence Livermore National Laboratory, 1987. submitted to the Journal of Computational Physics.

[215] J. B. Bell, P. Colella, and M. Welcome. Conservative front-tracking for inviscid compressible flow. In D. Kwak, editor, *Proceedings of the AIAA Tenth Computational Fluid Dynamics Conference*, pages 814-822, 1991. AIAA Paper 91-1599.

[216] H. Paillère, K. G. Powell, and D. De Zeeuw. A wave-model-based refinement criterion for adaptive-grid computation of compressible flows. In *30th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 1992. AIAA-92-0322.

[217] D. De Zeeuw and K. G. Powell. Euler calculations of axi-symmetric under-expander jets by an adaptive-refinement method. In *30th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 1992. AIAA-92-3521

[218] Y.-L. Chiang, B. van Leer, and K. G. Powell. Simulation of unsteady inviscid flow on an adaptively refined cartesian grid. In *30th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 1992. AIAA-92-0443.

[219] B. K. Edgar and P. r. Woodward. Diffraction of a shock wave by a wedge: Comparison of PPM simulations with experiment. In *30th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 1992. AIAA-92-0696.

[220] R. D. O'Dell and R. E. Alcouffe. Transport calculations for nuclear analyses: Theory and guidelines for effective use of transport codes. Technical Report LA-10983-MS, Los Alamos National Laboratory, 1987.

[221] E. E. Lewis and Jr. W. F. Miller. *Computational Methods of Neutron Transport*. Wiley-Interscience, 1984.

[222] B. Cockburn, S.-Y. Lin, and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws III: One-dimensional systems. *Journal of Computational Physics*, 84:90-113, 1989.

[223] G. Patnaik, R. H. Guirguis, J. P. Boris, and E. S. Oran. A barely implicit correction for flux-corrected transport. *Journal of Computational Physics*, 71:1-20, 1987.

[224] H. Akima. A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM*, 17:589-602, 1970.

[225] H. Akima. A method of bivariate interpolation and smooth surface fitting based on local procedures. *Communications of the ACM*, 17:18-31, 1974.

[226] H. Akima. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Transactions on Mathematical Software*, 4:148-159, 1978.

[227] H. Akima. On estimating partial derivatives for bivariate interpolation of scattered data. *Rocky Mountain Journal of Mathematics*, 14:41-52, 1984.

[228] P. L. Roe. Discrete models for the numerical analysis of time-dependent multidimensional gas dynamics. *Journal of Computational Physics*, 63:458-476, 1986.

[229] H. Öksüzoğlu. State vector splitting for the Euler equations of gasdynamics. In *30th Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 1992. AIAA-92-0326.

[230] S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27:919-940, 1990.

[231] B. Einfeldt, C. D. Munz, P. L. Roe, and B. Sjögreen. On Godunov-type methods near low densities. *Journal of Computational Physics*, 92:273-295, 1991.

[232] P. L. Roe. Modern numerical methods applicable to stellar pulsation. In J. R. Buchler, editor, *The Numerical Modelling of Nonlinear Stellar Pulsations*, pages 183-213, 1990.

[233] B. van Leer. Towards the ultimate conservative difference scheme. III. Upstream-centered finite-difference schemes for ideal compressible flow. *Journal of Computational Physics*, 23:263-275, 1977.

[234] P. Colella. Multidimensional upwind methods for hyperbolic conservation laws. *Journal of Computational Physics*, 87:171-200, 1990.

[235] H. Deconinck, P. L. Roe, and R. Struijs. A multidimensional generalization of Roe's flux difference splitter for the Euler equations. In H. A. Dwyer, editor, *Proceedings of the Fourth International Symposium on Computational Fluid Dynamics*, pages 282-287, 1991.

[236] C. Rumsey, B. van Leer, and P. L. Roe. Effect of a multi-dimensional flux function on the monotonicity of Euler and Navier-Stokes computations. In D. Kwak, editor, *Proceedings of the AIAA Tenth Computational Fluid Dynamics Conference*, pages 32-46, 1991. AIAA Paper 91-1530.

[237] D. Kontinos and D. McRae. An explicit, rotated upwind algorithm for the solution of the Euler/Navier-Stokes equations. In D. Kwak, editor, *Proceedings of the AIAA Tenth Computational Fluid Dynamics Conference*, pages 47-59, 1991. AIAA Paper 91-1531.

[238] J. K. Dukowicz and J. W. Kodis. Accurate conservative remapping (rezoning) for arbitrary Lagrangian-Eulerian computations. *SIAM Journal on Scientific and Statistical Computing*, 8:305-321, 1987.

[239] L. Fezoui and B. Stoufflet. A class of implicit upwind schemes for Euler simulations with unstructured meshes. *Journal of Computational Physics*, 84:174-296, 1989.

[240] G. Strang. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, 5:506-517, 1968.

[241] D. Gottlieb. Strang-type difference schemes for multidimensional problems. *SIAM Journal on Numerical Analysis*, 9:650-660, 1972.

[242] P. K. Smolarkiewicz. A fully multidimensional positive definite advection transport algorithm with small implicit diffusion. *Journal of Computational Physics*, 54:325-362, 1984

[243] P. L. Roe. Discontinuous solution to hyperbolic systems under operator splitting. *Num. Meth. Part. Diff. Eq.*, 7:277–297, 1991.

Curriculum Vita

William Jackson Rider was born in Spokane, Washington, September 5, 1963 to Carol Winslett Rider and Frank William Rider. After spending nine months in Spokane, he moved to Oberammergau Germany where his father was stationed in the United States Army. After three years in Germany, he returned to the United States with short stays in El Paso, Texas and Fort Sill, Oklahoma moved back to Spokane while his father spent a tour of duty in Vietnam. Upon his father's return to the United States, the Author moved back to Fort Sill (Lawton) where he spent the next six years attending elementary school. In 1975, his family moved to Germany once more, living two years in Aschaffenburg and Stuttgart each. After their return to the United States, the Author's family to Albuquerque, New Mexico where his father retired from the Army. The Author graduated from Eldorado High School in 1982.

He enrolled at the University of New Mexico in the Fall of 1982. On July 27, 1985, the Author married the former Felicia Anne Forbes. In the Spring of 1987, the Author graduated with a bachelor of science degree in nuclear engineering. He began his graduate work immediately following at the University of New Mexico. He completed his master's degree in the Summer of 1989. The Author's master thesis was entitled, "Parametric and Transient Analyses of the SP-100 System."

At that time the Author took a job at Los Alamos National Laboratory in the Reactor Design and Analysis Group, N-12. He has worked on a number of projects including: two-phase thermal-hydraulic code development, code development and analysis for gas-cooled terrestrial and extraterrestrial reactor systems, accelerator transmutation of waste, high-order solute tracking in reactor thermal-hydraulics and as well as other endeavors. He has presented papers at several conferences in the past several years. The Author has published a paper in the *AIAA Journal of Propulsion and Power*, and submitted a number of papers to various journals for review. The Author is a member of the American Institute for Aeronautics and Astronautics (AIAA) and the Society for Industrial and Applied Mathematics (SIAM).

How This Document Was Prepared

This document was prepared on a SPARCStation2 running SunOS 4.1; using LaTeX version 2.09 with TeX version 3.0.0. The BibTeX bibliography database version 0.99c was used to format the references. DVILASER/PS Sun Version 6.2.1 was used to produce the postscript output file, which exceeds 33 megabytes in size.

The line drawings were drawn on a Macintosh IIcx using a combination of Canvas version 2.0 and Aldus Freehand version 2.02. The two-dimensional plots were done with Kaleidagraph version 2.02. The three-dimensional plots of the slope limiters were done with Mathematica version 1.2. The surface and contour plots in Chapter F were drawn with CA-DISSPLA subroutines on a X-MP4/16 and transferred to a SPARCStation2 with the utility PPS as postscript files. All the Macintosh files were converted to postscript by Aldus Freehand version 2.02.

All of the computers mentioned above are located at Los Alamos National Laboratory.

U.S. Government Printing Office: 1992-671-036/60013